

Numerical Algorithms for use in a Dynamical Model of the Ocean

Alistair James Adcroft

A thesis submitted for the degree of

Doctor of Philosophy
University of London

and the

Diploma of Imperial College

Abstract

A new ocean circulation model based upon the Navier-Stokes equations is presented as a general purpose tool to study oceanic flows from the small scales of convective processes (~ 1 km), right through to the scale of the global circulation. The horizontal discretisation of the model is based upon the Arakawa ‘C’ grid, but includes a new method that overcomes the problem of spurious grid-scale noise that would otherwise be manifest at low resolutions (relative to the Rossby radius of deformation). The model is formulated in terms of finite volumes which allows an accurate representation of topography to be implemented.

Various classes of wave motion, inherent in the model physics, are derived by linearisation of the governing equations. The distinct wave motions are translated to the context of shallow water theory and analysed. A thorough understanding of how the grid-scale noise behaves in numerical models is developed. Two types of fundamental wave motion need to be modeled accurately; inertia-gravity waves and Rossby waves. The dispersive properties of these waves in a discretised model dictate how grid-scale noise propagates. A new numerical algorithm is then presented that evaluates the relevant terms in a more exact manner. The algorithm is shown to accurately model the dispersive properties of both inertia-gravity waves and Rossby waves.

Conventional representation of topography, as boxes fitted to the model grid, is severely limited by vertical resolution. The finite volume method, presented here, introduces the concept of zone (volume averaged) quantities and flux (area averaged) quantities into which the model variables can be categorised. The continuous governing equations are then integrated over finite volumes that fit the bottom topography, and are written explicitly in terms of zone and flux quantities. The model is thus able to resolve small variations in bottom relief without explicitly needing the equivalent vertical resolution in the interior.

Contents

Abstract	2
Table of Contents	3
List of Figures	5
List of Tables	7
Acknowledgments	8
1 Introduction	9
1.1 A historical perspective of numerical ocean modelling	9
1.2 Opportunities for a New Ocean Circulation Model	10
1.3 Development of a Navier-Stokes model for study of ocean circulation	10
1.4 Motivation for this thesis	11
1.5 Structure of this thesis	11
2 Equations of Oceanic Motion	13
2.1 Navier-Stokes equations of oceanic motions	13
2.2 Approximations	13
2.2.1 Acoustic modes	14
2.2.2 The Anelastic Approximation	15
2.2.3 Boussinesq approximation	16
2.2.4 Pressure equation (diagnostic)	16
2.3 Free surface	17
2.3.1 External gravity wave	18
2.3.2 Rigid-lid approximation	19
2.4 Classes of Motion	20
2.4.1 Non-hydrostatic Inertia-Gravity waves	20
2.4.2 Hydrostatic approximation: Hydrostatic Inertia-Gravity waves	22
2.4.3 Rossby Waves	22
2.5 Summary and comments	23
3 A Navier-Stokes Ocean Model	25
3.1 Finite difference methods	25
3.2 Finite difference notation and rules	27
3.3 Continuous formulation of model	27
3.4 Spatial discretisation of model	28
3.4.1 Conservative advection	30
3.5 The time-stepping scheme (Adams-Bashforth II)	31
3.6 Model Algorithm	32
3.7 3-D Elliptic Inversion	33
3.8 Computational Aspects of Model: Data parallel architectures	34
3.9 Applications of the model	35
3.9.1 Baroclinic instability of a chimney (convection site)	35
3.9.2 Climatological spin-up of the North Pacific Ocean	36
3.9.3 Conventional treatments of grid scale noise	39

4	Numerical Representation of Inertia-Gravity Waves	40
4.1	Shallow Water Theory	40
4.2	Inertia-Gravity Waves	43
4.3	Damped wave motion	43
4.4	Finite-differenced Inertia-gravity waves	44
4.5	Numerical Shallow Water models	48
4.5.1	B grid shallow water model	48
4.5.2	C grid shallow water model	49
4.5.3	Shallow water model results	49
4.6	The Explicit Oscillator Algorithm	54
4.6.1	Implicit Coriolis: damping of the Extra Inertial Mode	55
4.6.2	Comparison to the standard models	56
4.7	Summary	60
5	Numerical Representation of Rossby Waves and the C_d scheme	61
5.1	Rossby Waves	61
5.2	Finite-differenced Rossby waves	62
5.3	Time staggered grids	64
5.4	The C_d grid scheme: the CD hybrid	67
5.5	Determining the optimal coupling	69
5.6	Shallow water model results	70
5.7	Implementation of the C_d scheme in the GCM	76
5.8	Discussion	76
5.9	Summary	78
6	Representing Topography: the finite volume approach	79
6.1	Formulations in numerical modeling	80
6.2	Finite volume method	81
6.2.1	Continuity equation and boundary conditions	82
6.2.2	Tracers	82
6.2.3	Momentum Equations	84
6.2.4	Numerical stability	87
6.2.5	Comment on Accuracy	88
6.3	Testing the finite volume approach	88
6.3.1	Topographic β	88
6.3.2	Flow over a Gaussian bump	92
6.4	Conclusions	101
7	Concluding remarks	102
7.1	The C_d scheme	103
7.2	Shaved cells	103
7.3	Future development of model	104
7.4	Future directions in ocean modelling	104
A	Derivation of the Navier-Stokes Equations of Oceanic Motion	106
A.1	Conservation of Mass	106
A.2	Conservation of momentum	107
A.3	Conservation of salt	108
A.4	Thermodynamics (Continuity of heat, conservation of energy and potential temperature)	108
A.5	Pressure equation (prognostic)	110
B	Scaling of non-hydrostatic effects	111
C	Solution for zonal flow over a Gaussian bump	113
	Bibliography	115

List of Figures

2.1	Schematic of the single layer shallow water model	18
3.1	Describing $f(x)$ with discrete variables f_i	26
3.2	Three-dimensional ‘C’ grid	29
3.3	Flow diagram for the pre-conditioned conjugate gradient algorithm	34
3.4	Decomposition of ocean domain	34
3.5	Density anomaly at $z=-250\text{m}$ at day 10 in a convection experiment	35
3.6	Pressure (in m) at $z=-12.5\text{ m}$ at end of year 43.	37
3.7	Temperature (in $^{\circ}\text{C}$) at $\lambda=171.5\text{E}$ at end of year 43.	37
3.8	Salinity (in psu) at $\lambda=171.5\text{E}$ at end of year 43.	37
3.9	Vertical velocity at the base of the top layer after one month	38
3.10	Vertical velocity at $z=-3200\text{m}$ in the model after one month	38
4.1	Schematic of the two layer shallow water model	41
4.2	Arakawa grids A-E	44
4.3	Non-dimensional frequency of inertia-gravity waves	46
4.4	Non-dimensional frequency of discrete inertia-gravity waves at high resolution	46
4.5	Non-dimensional frequency of discrete inertia-gravity waves at low resolution	47
4.6	Low resolution ($4S/\Delta x^2 = 1/5$) integrations of the B grid (left) and C grid (right) shallow water models	50
4.7	Low resolution ($4S/\Delta x^2 = 1/5$) integrations of the B grid (left) and C grid (right) shallow water models	51
4.8	High resolution ($4S/\Delta x^2 = 5$) integrations of the B grid (left) and C grid (right) shallow water models	52
4.9	High resolution ($4S/\Delta x^2 = 5$) integrations of the B grid (left) and C grid (right) shallow water models	53
4.10	Low resolution ($4S/\Delta x^2 = 1/5$) integrations of the explicit oscillator scheme on the C grid	57
4.11	High resolution ($4S/\Delta x^2 = 5$) integrations of the explicit oscillator scheme on the C grid	58
4.12	Explicit Oscillator scheme at high resolution ($4S/\Delta x^2 = 1/2$) integrations on a β -plane	59
5.1	Rossby wave dispersion relations for the continuum and high resolution finite difference models	64
5.2	Rossby wave dispersion relations for the continuum and low resolution finite difference models	65
5.3	The Eliassen time horizontally staggered grids, AA’,BB’,CC’,DD’,EE’	66
5.4	The single step C grid compared to the time staggered grids DD’ CD	67
5.5	The two step C_d grid.	68
5.6	h , ζ and D for the B, C and C_d models at $t = 40$ in experiment I	72
5.7	h , ζ and D for the B, C and C_d models at $t = 5000$ in experiment I	73
5.8	h , ζ and D for the B, C and C_d models at $t = 40$ in experiment II	74
5.9	h , ζ and D for the B, C and C_d models at $t = 5000$ in experiment II	75
5.10	Vertical velocity at the base of the top layer in the model after one month	77
5.11	Vertical velocity at $z=-3200\text{m}$ in the model after one month	77
6.1	Four methods of representing topography in numerical models	80
6.2	A schematic of a shaved cell	81
6.3	Centres of volumes for shaved cells	84
6.4	Pressure and velocity vectors at $t = 1\text{ yr}$ for the flat bottomed β -plane integration.	90
6.5	Pressure and velocity vectors at $t = 1\text{ yr}$ for the sloping bottom f-plane integration.	91
6.6	Schematic flow over an isolated topographic feature.	92
6.7	Barotropic stream function at day 60 with constant zonal flow removed.	93
6.8	Barotropic stream function at day 60 with analytic solution removed.	93
6.9	Barotropic stream function Ψ from the SPEM integration	95

6.10	Barotropic stream function Ψ from the shaved cell integration	95
6.11	Depth Integrated Relative Vorticity in the SPEM integration	96
6.12	Depth Integrated Relative Vorticity in the shaved cell integration	96
6.13	Density anomaly at $z = -4000$ m in the SPEM integration	97
6.14	Potential Temperatures at $z = -4219$ m in the shaved cell integration	97
6.15	Schematic of Gaussian bump represented by “step-topography” and “shaved-cells”	98
6.16	Potential temperature at $z=-3656$ m and $t=1$ day using the step-wise representation of topography. . .	100
6.17	Potential temperature at $z=-3656$ m and $t=1$ day using the shaved cell representation of topography. .	100

List of Tables

3.1	GCM mixing and diffusion coefficients for the North Pacific spin-up	36
4.1	External parameters used in the comparison of the B and C grid shallow water models	49
5.1	Parameters for the double gyre integrations of the B grid, C grid and C_d models.	70
6.1	Parameters for the two experiments to model planetary (I) and topographic (II) β effects.	89
6.2	Parameters for the the comparison experiment to SPEM	94

Acknowledgments

I take this opportunity to thank my supervisor, John Marshall, for the advice he has given me throughout the last few years, and especially for the direction and opportunities he has given me by involving me in the model development.

I also wish to thank the other members of the development team; Chris Hill, Curtis Heisey and Lev Perelman. I especially thank Chris without whose attention to detail and organisation, the model would not have come as far as it has, and whose open-mindedness allowed me to try out many schemes in the model. I only wish more of them had worked. Thanks also to Dimitri Menemenlis for proof reading and for allowing me full reign over the chalk board.

I thank Paul Cloke who inadvertently interested me in the “gridding” issue and who, in the early days, schooled me in the art of numerical methods.

I am grateful to the Natural Environment Research Council for providing a studentship, to the Space and Atmospheric Physics group of Imperial College and to the faculty and staff of the Center for Meteorology and Physical Oceanography at MIT for hosting me in the recent years.

I would like to thank my examiners, Dr. Dan Moore and Dr. David Webb for their comments on an earlier revision of the text, and for their willingness to examine this thesis at such short notice.

And finally, I thank my fiancée, Sonya Legg, for her support, especially on those days when all seemed lost.

Chapter 1

Introduction

1.1 A historical perspective of numerical ocean modelling

To understand the ocean circulation and the role of the ocean in the climate system, we must understand complex processes in the ocean that occur on a wide range of scales. We use numerical models as tools for furthering our understanding of both the large scale circulation and small scale processes. The large scale circulation is intimately linked with processes on smaller scales. These processes must be either resolved or accurately parameterised if we are to render an accurate picture of the global ocean. The scale of geostrophic turbulence in the ocean is of the order 30–100 km which is very much smaller than the typical 1000 km scale in the atmosphere. The limitations on spatial resolution of numerical models, imposed by computer technology, are therefore more severe for ocean modelling than for meteorology.

Since the earliest attempts to use computers to help understand the ocean circulation computers have become ever more powerful. The computers accessible by most oceanographers today are orders of magnitude faster, and can store much more data, than those available in the 1960's when the first real efforts to model the ocean were being made. The earliest attempts to model the global scale circulation could not possibly resolve the meso-scale features present in the ocean. Instead, these processes had to be parameterised and, indeed, this is still the norm in regional and global calculations.

The first large scale ocean simulation was carried out by Sarkisyan [Sar55] in the mid '50s. Subsequent work by Sarkisyan focused on regional studies such as for the North Atlantic. Bryan began modelling the ocean at GFDL by applying numerical methods to the solution of the barotropic vorticity equation [Bry63] in the early '60s. Later, ocean modelling started elsewhere with Friedrich constructing a multi-level model in West Germany [Mos66]. Bryan, in the meantime, moved onto three-dimensional box models and started a number of multi-level primitive equation studies with Cox [BC68]. Soon after, a more general model was developed that incorporated irregular coast-lines and variable bottom topography [Bry69]. The methods laid out by Bryan and Cox have been the foundation of many subsequent modelling efforts by other parties.

The limitations imposed by the state of the computer technology meant that the Cox and Bryan model was first applied to regional simulations. These included regional simulations of the Southern, Indian and North Atlantic Oceans. Another model was then being developed in parallel at UCLA by Haney, Arakawa and Takano [Han71]. Takano went on to develop a less general model but applied it to the global ocean. He applied idealized atmospheric forcing and obtained an ocean circulation with many realistic features [Tak74, Tak75].

Since then, several multi-level primitive equation models have been developed. Semtners code [Sem74] is a derivative of the Bryan and Takano models. Two attempts to formulate the model on an Arakawa C grid, following the UCLA Atmospheric GCM were made. Jeong-Woo Kim [Kim79] and Cox (unpublished) found the models to be susceptible to grid-scale noise and they are apparently no longer in use [Jr.86]. Nevertheless, atmospheric models were being successfully built on C grids since they resolved the geostrophic adjustment process¹. Ocean models were constructed on the B grid since they were unable to resolve the Rossby radius of deformation which is much smaller in the ocean than in the atmosphere.

Further adaptations of the Bryan formulation were made by Han [Han84] and Cox [Cox84] and now, in the mid 1990's, there are many models derived from or based on this formulation. Recent development of the Bryan-Cox-Semtner code has been carried out at Los Alamos by Dukowicz and Smith [DSM93, DS94]. They have moved from a stream function approach for treating the barotropic mode to a surface pressure approach. This allows islands to

¹The inherent strengths and weaknesses of the B and C grid formulations as a function of resolution is described in detail in chapter 4.

be treated more appropriately.

All these models use height as the vertical co-ordinate. More recent models have departed from this convention. MICOM, developed by Smith and Bleck [SBB90], is an isopycnal model where height is a prognostic variable and potential density is the vertical co-ordinate. The advantage of isopycnal co-ordinates is that the parameterization of sub-grid scale processes can be made adiabatic. The formulation does however introduce complications both at the solid boundaries and at the surface, and has difficulties when incorporating an accurate equation of state. SPEM, see Haidvogel et al., 1991 [HWY91], and the Princeton model, see Mellor 1992 [Mel92], use terrain following σ co-ordinates. These models are well suited for coastal oceanography where high horizontal resolution allows σ co-ordinates to follow the topography smoothly. The SPEM code has not been widely applied at global scales presumably because the topography of the ocean is extremely irregular and has many islands. The Princeton model uses a variant of σ co-ordinates where the number of modes in the vertical can be varied in the horizontal to allow sudden changes in topography.

1.2 Opportunities for a New Ocean Circulation Model

The advent of new parallel computer architectures has recently allowed oceanographers to begin running models at eddy resolving resolutions on a global scale (see, for example Semtner, 1988 [SC88]). As the available memory of computers increases, issues involving grid resolution must be re-addressed. One such issue concerns the approximations made to the governing equations used for large scale ocean modelling. For example, the usual approximation of hydrostatic balance in the vertical may not be strictly valid at the smaller scales (smaller than the Rossby deformation radius) now being resolved by these models. Further, horizontal Coriolis effects can only be investigated if the hydrostatic approximation is relaxed. Small scale phenomena are interesting in their own right and can only be investigated using a non-hydrostatic model. Currently, all global ocean models are hydrostatic and there has not been, until now, a numerical model applicable to both the small and the large scale. Such a model would allow a smooth transition between studies at high and low resolution.

Models which resolve the Rossby radius of deformation operate in a parameter regime analogous to that of most atmospheric models. They do not, therefore, suffer from the excessive grid scale noise, confronted independently by Kim and by Cox.

Just as advances in computers allow the global circulation models to increase their resolution, the same advances permit models designed primarily for the study of small scale phenomena to be applied to the larger scale.

Here, one such model is described that is applicable to all scales in the ocean. It has successfully been applied to the convective overturning scale, up through the meso-scale and in extended integrations at the global scale. The model is non-hydrostatic, though it can operate in a hydrostatic or quasi-hydrostatic mode. It is formulated on a C grid, but uses an innovative method for evaluating the Coriolis term so that it is not susceptible to the grid-scale noise problems of C grid models.

1.3 Development of a Navier-Stokes model for study of ocean circulation

The model described here uses height as a vertical co-ordinate and can have arbitrary topography and irregular coast-lines (or may be periodic in x and/or y). The kernel of the model is founded on the incompressible Navier-Stokes equations. It can also be used to step forward quasi-hydrostatic and hydrostatic models that employ approximated forms of the governing equations.

The model incorporates ideas developed in the computational fluid dynamics community which are relatively new to ocean modelling (eg. conjugate gradient methods and finite volume methods). It has been developed on a parallel computer architecture that gives the modeller access to higher resolution, through increased memory, and to longer integrations, through increased speed.

The model solves the incompressible Navier-Stokes equations and so involves fewer assumptions than the hydrostatic primitive equations employed in most existing models. Further approximated forms can be recovered by means of “switches” so that the relative importance of various small terms can be evaluated. In particular, the non-hydrostatic facility can be turned on or off selectively; a quasi-hydrostatic form of the model allows horizontal Coriolis effects to be retained whilst neglecting the advection of vertical momentum.

The Navier-Stokes model has been designed for the study of dynamical processes in the ocean ranging from the convective scale, through the geostrophic eddy scale to the global scale circulation. The algorithm for studying

this wide range of scales is essentially the same, though approximations² can be made at larger scales to make the integration more efficient with no significant loss of accuracy.

The computational challenge is to maintain non-divergence of the flow. This entails diagnosing the appropriate pressure field that ensures the flow has zero divergence at all times. The equation satisfied by the pressure field is elliptic and has Neumann boundary conditions. Thus, the pressure field depends upon the global distribution of inhomogeneous sources and boundary conditions.

This last aspect of the problem has implications for implementation on parallel computers. Global interaction, required by the elliptic problem, demands that information be exchanged globally between processors on a parallel machine. The reduction of inter-processor communication is thus a priority in designing such an algorithm.

The development of the model has required expertise from diverse fields. Knowledge of ocean physics, numerical methods, parallel computer architectures, data management and visualisation were supplied by members of a large team; nominally John Marshall, Chris Hill, Lev Perelman, Curtis Heisey, and myself. Discussions with Prof. Arvind, Andrew White, Roger Brugge, and Paul Cloke, among others, helped guide the development of the model.

1.4 Motivation for this thesis

The model is developed on a ‘C’ grid because it is the appropriate grid for study of small scale phenomena. It is also the natural grid for both a finite volume formulation and the pressure correction method. When applied at coarse resolution, the pure ‘C’ grid model is found to be susceptible to grid-scale noise, as was the case for Kim and for Cox. This is a direct consequence of the choice of gridding and is a well documented problem, most notably by Arakawa and Lamb [AL77]. To evaluate the Coriolis term on the C grid, the horizontal velocities must be spatially averaged. A result of the spatial averaging is that the Coriolis term vanishes for the grid scale. Grid scale noise can therefore exist as stationary waves that have no inertial oscillatory component. For this reason, low resolution ocean models are formulated on other grids, typically the Arakawa B grid.

This thesis has two foci. The first concerns the representation of the Coriolis term on a staggered grid, or, more generally, on the representation of inertia-gravity and Rossby waves in numerical models. Noise, manifest in the ‘w’ field, was resilient to many attempts to control it. Many approaches, ranging from brute-force filtering to the introduction of artificial damping terms, were attempted. None of these methods were satisfactory for various reasons. We decided, therefore, to find a correction to the problem rather than try to control it. This led to the formulation of what we term the C_d scheme; a D grid is used in tandem with the C grid where the D grid velocities are used to evaluate the Coriolis term. The scheme can be successfully used at coarse resolution and avoids the grid scale noise problems that would otherwise manifest themselves.

The second issue addressed in this thesis is the representation of topography. The conventional representation of topography in height co-ordinate models is as “step-wise” functions fitted to the model layer depths, a crude representation. As an alternative, we consider a finite volume approach in which shaved cells can be used to represent topography. The ‘C’ grid formulation lends itself quite naturally to a finite volume interpretation of the model. The finite volume approach aims to conserve properties such as volume and tracers in a precise manner. The model equations are discretised by integrating them over a grid of finite volumes. In the interior, the use of regular volumes or cells gives rise to a conventional discretisation. However, the cells need not be regular where they abut a solid boundary. We take advantage of this by shaving the cells to fit the topography of the ocean. In this manner, the model is able to represent topographic effects that could otherwise not be represented without prohibitive increases in horizontal and vertical resolution.

1.5 Structure of this thesis

Chapter 2 presents the continuous equations of motion on which the model is based. The approximations implicit in the Navier-Stokes equations are discussed. Finally the natural modes of motion at various degrees of approximation are derived. Subsequent chapters concern the accurate representation of these motions in numerical models.

The kernel of the numerical model, excluding the implementation of the two innovations, is described in chapter 3. Here, the model is integrated in a realistic configuration and found to be susceptible to grid-scale noise when the Rossby radius of deformation is not resolved. The nature of the noise is traced to the spatial averaging of the Coriolis term on a C grid. Several attempts to control the noise problem are briefly outlined and discussed.

Chapters 4 and 5 deal with the issue of the spatially averaged Coriolis term on a C grid. The similarity between the shallow water equations of motion and the equations pertaining to a baroclinic mode in the Navier-Stokes model

²The hydrostatic approximation is described at the end of chapter 2.

is used to motivate a study of the gridding issue in the context of the shallow water equations. Chapter 4 reviews the work of Arakawa and Lamb [AL77] and describes a scheme that improves the representation of inertia-gravity waves. The scheme is flawed due its inability to represent Rossby waves. A new scheme, the C_d scheme, is then described in chapter 5, which correctly treats both inertia-gravity waves and Rossby waves.

The C_d scheme is implemented in a numerical shallow water model and compared with the B and C grids. It is found to have no grid scale noise problems and is subsequently implemented in the Navier-Stokes model. The scheme entails only minor modification of the original model (as described in chapter 3) and adds little computational overhead.

In chapter 6, the representation of topography is discussed. The finite volume approach and the use of shaved cells to represent topography are described. The model is re-formulated using finite volumes. A series of experiments are devised and conducted to test the representation of topography using shaved cells.

Chapter 2

Equations of Oceanic Motion

All classes and scales of motion are described by the Navier-Stokes equations (derived in Appendix A). The equations used for the study of the general circulation of the ocean, however, are based on approximated forms which involve the Boussinesq approximation and assume non-divergence of the flow. The latter approximation excludes the acoustic modes of motion.

Further degrees of approximation are made that modify but do not eliminate the remaining natural modes of motion. These classes of motion will be derived. The accurate representation of these modes was a paramount concern when building and understanding the finite difference model. Chapters 4 and 5 deal exclusively with the numerical representation of inertia-gravity waves and Rossby waves, the nature of which will be derived at the end of this chapter.

2.1 Navier-Stokes equations of oceanic motions

The complete unapproximated system that describes inviscid, adiabatic flow is:

$$\frac{Dp}{Dt} + \rho c_s^2 \nabla \cdot \mathbf{u} = 0 \quad (2.1a)$$

$$\frac{D\mathbf{u}}{Dt} + 2\boldsymbol{\Omega} \wedge \mathbf{u} + \frac{\nabla p}{\rho} + \nabla \Phi = 0 \quad (2.1b)$$

$$\frac{DS}{Dt} = 0 \quad (2.1c)$$

$$\frac{D\theta}{Dt} = 0 \quad (2.1d)$$

$$\rho = \rho(\theta, S, p) \quad (2.1e)$$

where p , $\mathbf{u} = [u, v, w]$, θ , S and ρ are the pressure, three-dimensional velocity, potential temperature, salinity and in-situ density respectively. These equations are derived in Appendix A from first principles.

There are six prognostic equations for the variables p , \mathbf{u} , θ and S and one diagnostic relation for ρ . These are the *Navier-Stokes* equations supplemented by complete thermodynamics. These equations are the basis from which all models of the ocean are derived in their various degrees of approximation.

The full system is often also expressed in terms of in-situ temperature T rather than potential temperature θ . Similarly, the salt and temperature equations can be combined into a thermodynamic equation for potential density defined $\sigma_{p_o} = \rho(\theta, S, p_o) \Rightarrow \frac{d}{dt}\sigma_{p_o} = 0$.

The pressure equation will later reduce to the continuity equation (expressing non-divergence of the flow) when the acoustic modes have been filtered out of the system. The pressure equation is obtained by combining conservation of mass with the equation of state. The Navier-Stokes equations could equally have been written using conservation of mass instead of the pressure equation. This, however, would have yielded two explicit equations for the density, ρ , and none for the pressure, p .

2.2 Approximations

System 2.1 is completely general and describes all physical processes in the ocean that are not affected by molecular viscosity and diffusivity. This, however, is a draw back for any practical computation since the explicit time scales

in the system require that all processes be resolved. For example, the acoustic modes are very fast relative to any process relevant to the long time scales of interest (normally much longer than a few days). An explicit numerical model based on the above system would have a time step limited by the sound waves. This severely limits the applicability of any such numerical model.

There are two ways of proceeding; either writing the numerical model in an implicit manner or explicitly filtering out the fast modes. Implicit models can be written, and in fact would bear a striking resemblance to the adjusted or filtered model. Implicit techniques act to slow the respective process down so that a long time step does not violate any criteria. The filtered system, in the other limit, assumes that the process acts infinitely fast so that the system is always adjusted. The latter approach is more common because it is considerably easier to implement.

Three types of approximation will be used throughout the rest of this chapter:

Rapid time scale approximation in which the mode under consideration is considered much faster than the time scales of interest. Filtering of these modes assumes that the process has acted infinitely fast and that the fluid is instantaneously adjusted. The acoustic modes and surface gravity waves will be filtered in this manner.

Short spatial scale approximation in which the spatial scale of a mode is very much shorter than any scale of interest. For instance, internal inertia-gravity waves behave in a hydrostatic manner except where the aspect ratio of the motion is small.

Small amplitude approximation in which the amplitude of an effect is much smaller than the signals of interest. For example, the amplitude of density perturbations due to the transit of an acoustic wave are typically much smaller than the dynamically interesting variations arising from mean vertical motion excursions.

These classifications are not exclusive and often one is implied by another.

In the account that follows, classes of motion will be derived and approximations made that either modify or filter the modes of motion.

2.2.1 Acoustic modes

To derive the acoustic modes of motion, consider a reduced form of the Navier-Stokes equations; the conservation of mass, momentum equations retaining just the local acceleration and pressure gradient terms and equation of state (the dependence on temperature and salinity can be neglected for adiabatic motion):

$$\frac{\partial}{\partial t} \mathbf{u} + \frac{1}{\rho} \nabla p = 0 \quad (2.2a)$$

$$\frac{\partial}{\partial t} \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (2.2b)$$

$$\rho = \rho(p) \quad (2.2c)$$

Combining the conservation of mass and the reduced equation of state gives the pressure equation. Linearizing about some mean state then yields:

$$\frac{\partial}{\partial t} \mathbf{u} + \frac{1}{\rho_o} \nabla p = 0$$

$$\frac{\partial}{\partial t} p + \rho_o c_s^2 \nabla \cdot \mathbf{u} = 0$$

The three dimensional divergence of the momentum equations is:

$$\frac{\partial}{\partial t} \nabla \cdot \mathbf{u} + \frac{1}{\rho_o} \nabla^2 p \quad (2.3)$$

This can be substituted into the time derivative of the pressure equation to yield:

$$\frac{\partial^2}{\partial t^2} p = c_s^2 \nabla^2 p \quad (2.4)$$

which is a non-dispersive wave equation describing waves with phase and group speed c_s . The speed of sound in water is approximately 1500 m s^{-1}

The acoustic modes act to adjust the pressure field so that the tendency for the three-dimensional divergence of the flow is to vanish. The wave motion would be eliminated if either of the time derivatives in 2.2a and 2.2b were removed, or if the pressure dependence of density were removed. Setting $\left. \frac{\partial \rho}{\partial p} \right|_{\theta, S} \rightarrow 0$ means that the speed of sound becomes infinite, $c_s \rightarrow \infty$.

2.2.2 The Anelastic Approximation

As mentioned earlier, filtering out of the sound waves can be achieved by assuming the limiting case of incompressibility that makes the sound speed become infinite, $c_s \rightarrow \infty$. An intermediate approximation, the anelastic approximation, that retains compressibility effects requires a reference state (denoted by the subscript r) to be defined as follows:

$$\begin{aligned} \frac{\partial p_r}{\partial z} &= -g\rho_r(z) \\ p &= p_r(z) + p' \quad \rho = \rho_r(z) + \rho' \end{aligned} \quad (2.5)$$

where both ρ_r and p_r are prescribed functions of the vertical co-ordinate. Deviations from the reference state are denoted by primes. The momentum equation is unapproximated.

$$\frac{\partial}{\partial t}\rho\mathbf{u} + \nabla \cdot (\rho\mathbf{u}\mathbf{u}) + 2\Omega \wedge \rho\mathbf{u} + \nabla p' + \rho'\nabla\Phi \quad (2.6)$$

The pressure equation becomes:

$$\frac{Dp_r}{Dt} + \frac{Dp'}{Dt} = -\rho c_s^2 \nabla \cdot \mathbf{u} \quad (2.7)$$

The anelastic approximation can be obtained by assuming that the perturbations in the pressure field propagate so fast that the pressure field is always adjusted. This means that $\frac{Dp'}{Dt} \ll \frac{Dp_r}{Dt}$ is a good approximation for the slow and long scales of interest. Noting that $\frac{Dp_r}{Dt} = w\frac{\partial p_r}{\partial z}$, the pressure equation can be approximated:

$$\nabla \cdot \mathbf{u} = \frac{g\rho_r}{c_s^2\rho} w \quad (2.8)$$

This assumption only deals with time scales and so the equation of state still has a dependence on p' . This is inconsistent with the continuity equation because there would then be two different prognostic equations for density. Therefore, a more consistent method is to, instead, make an assumption about the amplitudes of motion.

Assume that the amplitudes of motion are such that density changes induced by the acoustic pressure perturbations are very much smaller than the density changes brought about by large changes in depth (*i.e.* $\frac{\partial \rho}{\partial p'} \ll \frac{\partial \rho}{\partial p_r}$). Then the equation of state becomes:

$$\rho = \rho(\theta, S, p_r) \quad (2.9)$$

Differentiating this equation and making use of the continuity equation, the same anelastic continuity equation is derived. In this manner, the equation of state is consistent with the approximated pressure equation.

The complete anelastic system is then:

$$\nabla \cdot \mathbf{u} = \frac{g\rho_r}{c_s^2\rho} w \quad (2.10a)$$

$$\frac{D\mathbf{u}}{Dt} + 2\Omega \wedge \mathbf{u} + \frac{\nabla p'}{\rho} + \frac{\rho'}{\rho} \nabla\Phi = 0 \quad (2.10b)$$

$$\frac{DS}{Dt} = 0 \quad (2.10c)$$

$$\frac{D\theta}{Dt} = 0 \quad (2.10d)$$

$$\rho = \rho(\theta, S, p_r) \quad (2.10e)$$

It should be pointed out that there is no explicit equation for the pressure field p' . A diagnostic equation for p' can be deduced by combining the momentum equations with the anelastic continuity equation to form an elliptic equation. This will be done later.

The system no longer contains sound waves but does still contain the effects of compressibility brought about by the slow motions that involve changes in depth.

The removal of the acoustic modes reduces the number of natural modes to four. The appropriate number of prognostic equations is obtained by replacing the w equation with the anelastic equation so that there are three diagnostic equations; an elliptic equation for pressure, the anelastic equation and the equation of state.

2.2.3 Boussinesq approximation

The name *Boussinesq approximation* is not always used in a consistent fashion. In fact it refers to a wide range of very different degrees of approximation from anelastic flow, through non-divergent yet compressible flow, to fully incompressible flow. Two steps will be described here; the first makes use of an observation about the significance of density perturbations, the second notes that the scale height over which compressibility effects are important is very much greater than the real depth of the ocean.

In the ocean, variations in density are small compared to a typical mean value, $\rho'/\rho_o \approx 10^{-3}$. Thus, wherever the full density is used, it can be approximated by the mean value ρ_o . This is a trivial exercise when applied to the anelastic equations above. If the exercise is applied to the Navier-Stokes equations, special care is needed when considering the gravitational term. By taking out the reference pressure (now defined $\frac{\partial}{\partial z} p_r = -g\rho_r = -g\rho_o$ so that simply $p_r = -g\rho_o z$), the gravitational term is naturally handled.

The second step notes, that in the anelastic continuity equation, the two terms involving w are $\frac{\partial w}{\partial z}$ and $\frac{g}{c_s^2} w$. The scaling of these two terms goes like $\frac{1}{H}$ and $\frac{g}{c_s^2}$. The exponential scale depth $\frac{c_s^2}{g} \approx 225 \text{ km}$ is around 50 times larger than the deeper parts of the real ocean. Thus the anelastic term is typically quite small and can be neglected leaving the non-divergence condition.

The Boussinesq equations of motion for non-divergent yet compressible flow are:

$$\begin{aligned} \nabla \cdot \mathbf{u} &= 0 & (2.11a) \\ \frac{D\mathbf{u}}{Dt} + 2\Omega \wedge \mathbf{u} + \frac{\nabla p'}{\rho_o} + \frac{\rho'}{\rho_o} \nabla \Phi &= 0 & (2.11b) \\ \frac{DS}{Dt} &= 0 & (2.11c) \\ \frac{D\theta}{Dt} &= 0 & (2.11d) \\ \rho &= \rho(\theta, S, -g\rho_o z) & (2.11e) \end{aligned}$$

Note that the Lagrangian form of the equations is readily interchanged with the Eulerian or flux divergence from because of the non-divergence of the flow; $\frac{\partial}{\partial t} \phi + \nabla \cdot (\phi \mathbf{u}) = \frac{D\phi}{Dt}$. This is useful when formulating finite difference models since conservation is better expressed as the divergence of a flux.

There is now apparently an inconsistency between the equation of state which admits compressible effects and the statement of the non-divergent flow. Taking the time derivative of the equation of state:

$$\frac{d\rho}{dt} = \frac{1}{c_s^2} \frac{dp_r}{dt} = \frac{-g\rho_o}{c_s^2} w \quad (2.12)$$

If the continuity equation is still satisfied then the non-divergence of the flow would be violated:

$$c_s^2 \nabla \cdot \mathbf{u} = -gw \quad (2.13)$$

Therefore, the Boussinesq approximation must relax the principle of conservation of mass and instead replace this principle with the non-divergence condition. This equation is usually termed the *incompressibility condition* though this is misleading since the compressible effects felt through the dependence of temperature and density on depth are still incorporated. The incompressibility condition expresses a constancy of volume and is thus a pivotal equation for the formulation of a model built around fixed, finite volumes. Conservation of total mass has therefore been exchanged for constant total volume.

2.2.4 Pressure equation (diagnostic)

As mentioned earlier for the anelastic equations, there is no prognostic equation for the pressure in the Boussinesq equations, as written above. A prognostic equation for p' can be found by combining the incompressibility condition with the momentum equations. For convenience, first define the vector \mathbf{G} to be all the terms of the momentum equations except the local derivative and pressure gradient:

$$\mathbf{G} \equiv -(\mathbf{u} \cdot \nabla) \mathbf{u} - 2\Omega \wedge \mathbf{u} - \frac{\rho'}{\rho_o} \nabla \Phi \quad (2.14)$$

Now the momentum equations can be written simply:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\rho_o} \nabla p' = \mathbf{G} \quad (2.15)$$

The flow must be non-divergent for all time. Substituting the momentum equations into the incompressibility condition yields:

$$\frac{1}{\rho_o} \nabla^2 p' = \nabla \cdot \mathbf{G} - \frac{\partial}{\partial t} \nabla \cdot \mathbf{u} \quad (2.16)$$

where the last term should vanish. In practice, the last term is kept to stabilise the numerical model so that any divergence of the flow is seen by the pressure field and adjusted for.

The final form for the Boussinesq equations is:

$$\nabla^2 p' = \rho_o \nabla \cdot \mathbf{G} \quad (2.17a)$$

$$\frac{\partial \mathbf{u}_h}{\partial t} = \mathbf{G}_h - \frac{1}{\rho_o} \nabla_h p' \quad (2.17b)$$

$$\frac{\partial w}{\partial z} = -\nabla_h \cdot \mathbf{u}_h \quad (2.17c)$$

$$\frac{Dt}{Dt} \theta = 0 \quad (2.17d)$$

$$\frac{Dt}{Dt} S = 0 \quad (2.17e)$$

$$\rho = \rho(\theta, S, -g\rho_o z) \quad (2.17f)$$

$$\mathbf{G} = -(\mathbf{u} \cdot \nabla) \mathbf{u} - 2\Omega \wedge \mathbf{u} - \frac{\rho'}{\rho_o} \nabla \Phi \quad (2.17g)$$

The boundary conditions for the elliptic problem take the form of a Neumann condition on the normal gradient at the boundaries. The exact details pertaining to solid boundaries will be left until the numerical sections. The boundary condition applied at the free-surface is one of continuity of pressure across the interface, *i.e.* ocean surface pressure and atmospheric surface pressure will be the same.

2.3 Free surface

The surface of the ocean is free to move in response to the net accumulation of depth integrated mass fluxes and to mass fluxes across the interface (precipitation, evaporation, river discharge, freezing and thawing). The latter sources are normally grouped into the thermodynamics of the model. In the incompressible model, the mass flux is replaced by a volume flux, or, in other words, the continuity equation becomes the incompressibility condition. The equation of evolution for the free surface, ignoring sources and sinks of mass, can be obtained by considering the incompressibility condition, integrated over the total depth of the ocean:

$$\int_{-H(x,y)}^{h(x,y,t)} (\nabla_h \cdot \mathbf{u}_h + \frac{\partial w}{\partial z}) dz = \int_{-H(x,y)}^{h(x,y,t)} \nabla_h \cdot \mathbf{u}_h dz + [w(z)]_{-H(x,y)}^{h(x,y,t)} = 0 \quad (2.18)$$

where $z = -H(x, y)$ defines the solid ocean bottom and $z = h(x, y, t)$ defines the position of the air-sea interface. The Leibniz formula connects the depth integral of the horizontal divergence to the horizontal divergence of the barotropic flow:

$$\nabla_h \cdot \int_{-H(x,y)}^{h(x,y,t)} \mathbf{u}_h dz = \int_{-H(x,y)}^{h(x,y,t)} \nabla_h \cdot \mathbf{u}_h dz + (\mathbf{u}_h(z=h) \cdot \nabla h - \mathbf{u}_h(z=-H) \cdot \nabla(-H)) \quad (2.19)$$

The kinematic boundary condition applicable at both interfaces is that a particle on the interface will remain on the interface; $\frac{Dt}{Dt} z_{boundary} = w$. Applied to the free surface and bottom:

$$\text{Free surface: } \frac{Dt}{Dt} h - w_h = \frac{\partial h}{\partial t} + \mathbf{u}_h(z=h) \cdot \nabla h - w_h = 0 \quad (2.20)$$

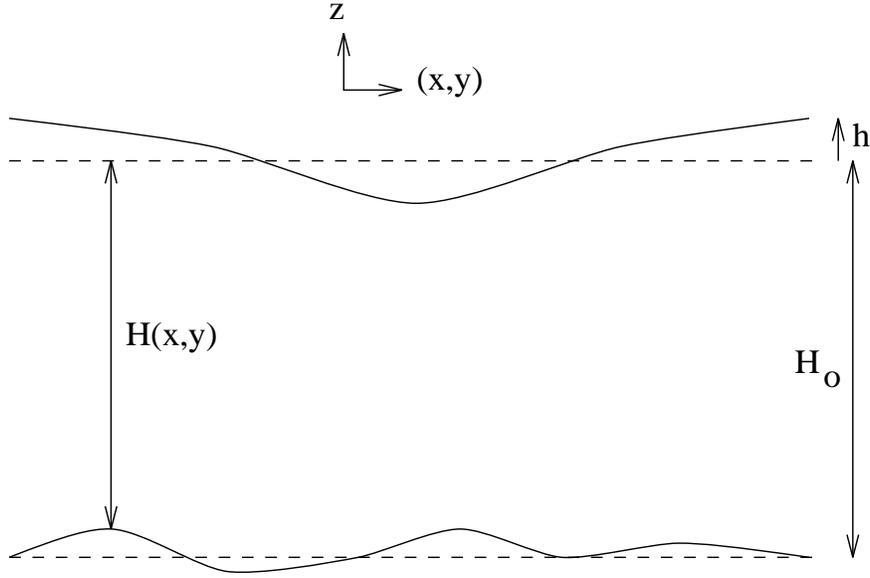


Figure 2.1: The shallow water model comprises a homogeneous layer of fluid between rigid bottom, of depth H , and free surface of, elevation h .

$$\text{Solid bottom: } \frac{Dt}{Dt}(-H) + w_H = \mathbf{u}_h(z = -H) \cdot \nabla(-H) - w_H = 0 \quad (2.21)$$

Defining the depth integrated flow:

$$\hat{\mathbf{u}}_h(x, y, t) \equiv \frac{1}{h + H} \int_{-H(x,y)}^{h(x,y,t)} \mathbf{u}_h dz \quad (2.22)$$

and substituting the boundary conditions into the the depth integrated continuity and making use of the Leibniz formula, the equation for surface elevation is obtained:

$$\frac{\partial h}{\partial t} + \nabla \cdot \{(h + H)\hat{\mathbf{u}}_h\} = 0 \quad (2.23)$$

The pressure is continuous across the free interface and so:

$$p'(x, y, z = h(x, y, t)) = p_a \quad (2.24)$$

where p_a is the atmospheric pressure, usually assumed small and constant, trivializing the boundary condition.

2.3.1 External gravity wave

Associated with the free surface is a relatively fast wave motion. Perturbations in the surface elevation tend to want to slump under gravity. The motion generated is then subject to the effects of the rotating frame of reference. Just as the acoustic modes were filtered, the external gravity wave will also be filtered later.

Consider an approximation of the ocean as a layer of homogeneous fluid but with a free surface (figure 2.1). Retaining only linear terms, the equations of motion for the layer are:

$$\frac{\partial u}{\partial t} - f_o v + \frac{1}{\rho} \frac{\partial p}{\partial x} = 0 \quad (2.25a)$$

$$\frac{\partial v}{\partial t} + f_o u + \frac{1}{\rho} \frac{\partial p}{\partial y} = 0 \quad (2.25b)$$

$$g\rho + \frac{\partial p}{\partial z} = 0 \quad (2.25c)$$

$$\frac{\partial h}{\partial t} + \nabla \cdot H\hat{\mathbf{u}}_h = 0 \quad (2.25d)$$

Integrating the hydrostatic balance, the homogeneity of the layer dictates that the horizontal pressure gradients be independent of depth:

$$p(x, y, z) = g\rho(h - z) + p|_{z=h} \quad (2.26)$$

$$\Rightarrow \nabla_h p = g\rho \nabla_h h \quad (2.27)$$

where the horizontal variations in atmospheric pressure have been neglected. The horizontal flow, \mathbf{u}_h , is then easily approximated as being barotropic also (a consequence of the Taylor-Proudman theorem).

For the flat bottom case (H is constant), linear barotropic equations can be written exclusively in terms of the barotropic flow and the free surface elevation:

$$\frac{\partial u}{\partial t} - f_o v + g \frac{\partial h}{\partial x} = 0 \quad (2.28a)$$

$$\frac{\partial v}{\partial t} + f_o u + g \frac{\partial h}{\partial y} = 0 \quad (2.28b)$$

$$\frac{\partial h}{\partial t} + H \nabla_h \cdot \mathbf{u}_h = 0 \quad (2.28c)$$

The system can be expressed in terms of the horizontal divergence, $D = \nabla_h \cdot \mathbf{u}_h$, and vertical component of vorticity, $\zeta = \hat{\mathbf{k}} \cdot (\nabla_h \wedge \mathbf{u}_h)$:

$$\frac{\partial D}{\partial t} - f_o \zeta + g \frac{\partial^2 h}{\partial x^2} = 0 \quad (2.29a)$$

$$\frac{\partial \zeta}{\partial t} + f_o D = 0 \quad (2.29b)$$

$$\frac{\partial h}{\partial t} + HD = 0 \quad (2.29c)$$

Taking the time derivative of the divergence equation and substituting in from the vorticity and height equations, the surface gravity wave equation is obtained:

$$\frac{\partial^2 D}{\partial t^2} + f_o^2 D - gH \nabla_h^2 D = 0 \quad (2.30)$$

The dispersion relation for plain waves of the form $\exp i(kx + ly - \omega t)$ is:

$$\omega^2 = f_o^2 + gH(k^2 + l^2) \quad (2.31)$$

For long waves ($|\hat{\mathbf{k}}|^2 \ll \frac{gH}{f_o^2}$) the frequency is almost constant corresponding to the inertial frequency. Short waves ($|\hat{\mathbf{k}}|^2 \gg \frac{gH}{f_o^2}$) are almost non-dispersive: $\omega \approx \sqrt{gH} |\hat{\mathbf{k}}|$. The phase and group speeds for such short waves is $c_\phi = c_g = \sqrt{gH}$ which for an 4 km deep ocean gives a wave speed of 200 m s⁻¹. Compared to other wave and fluid motions in the ocean, this is extremely fast.

2.3.2 Rigid-lid approximation

The fast surface gravity wave speed demands a strict limit on the possible time step allowed in the model (if the process is explicitly represented). For a 1 degree horizontal resolution model at mid-latitudes, a surface gravity wave speed of 200 m s⁻¹ would take only 500 s to travel across a grid-cell. The explicit time step would have to be some fraction of this. This is far short of that preferred to investigate the longer time-scales of interest.

The surface gravity wave can be filtered out of the system (just as the acoustic modes were) by imposing a rigid-lid on the model. The lid can be thought of as exerting a surface pressure on the model equal to the hydrostatic weight of the water column above the mean surface.

The upper boundary condition now becomes $w = 0$ (the surface is fixed) so that the vertical integral of the continuity (incompressibility) equation simply becomes:

$$\nabla_h \cdot \left(\int_{-H(x,y)}^{z=0} \mathbf{u}_h dz \right) = \nabla_h \cdot H \hat{\mathbf{u}}_h = 0 \quad (2.32)$$

where $\hat{\mathbf{u}}_h$ is the depth averaged (or barotropic) flow.

The full pressure field is split into two parts; a surface pressure, $p_s(x, y)$, representing the pressure exerted by the rigid lid, and the remaining internal pressure, p_i .

The horizontal momentum equations are written succinctly:

$$\frac{\partial u}{\partial t} + \frac{1}{\rho_o} \frac{\partial p_s}{\partial x} = G_u \quad (2.33a)$$

$$\frac{\partial v}{\partial t} + \frac{1}{\rho_o} \frac{\partial p_s}{\partial y} = G_v \quad (2.33b)$$

where all other terms, including the remaining pressure gradients are collected into the G terms. The divergence of the depth integrated horizontal momentum equations is:

$$\frac{1}{\rho_o} \nabla_h \cdot H \nabla_h p_s = \hat{G} - \frac{\partial}{\partial t} \nabla_h \cdot H \hat{\mathbf{u}}_h \quad (2.34)$$

where the last term vanishes through equation 2.32.

2.4 Classes of Motion

The Boussinesq approximation has removed a certain amount of non-linearity from the system. The natural modes of motion remain essentially unchanged (except that the acoustic modes have been filtered out of the system). The remaining four natural modes of motion are:

- 1× **Temperature-Salinity (T - S) mode** This mode exists because there are two active tracers, θ and S , that are dynamically felt through the density. Non-linearities in the equation of state can then lead to *cabbeling*. Differing mixing coefficients for θ and S allows *double diffusion*. Most importantly for the large scale circulation, the different nature of boundary conditions for θ and S leads to very complex behavior including the existence of multiple steady states.
- 2× **Gravity modes** As will be seen, external and internal gravity waves are modified extensively by rotation to produce *inertia-gravity* waves. Compressibility is an insignificant effect. These waves of horizontal divergence propagate energy quickly and bring the ocean into a geostrophically adjusted state.
- 1× **Geostrophic mode** This is the slowest of the natural modes but is perhaps the most important for shaping the large scale circulation. Asymmetry in the dispersive properties of these waves leads to east-west asymmetry of the oceans.

For the remainder of this chapter, the non-linearities in the equation of state will be neglected for convenience. Apart from the effects of mixing and boundary conditions, this trivializes the roles of the two thermodynamic tracers θ and S and allows them and the equation of state to be replaced by a prognostic equation for density:

$$\frac{d\rho}{dt} = 0 \quad (2.35)$$

which has no connection with the continuity equation. Under this assumption, only the gravity modes and the geostrophic mode will be apparent.

Both the gravity and the geostrophic classes of motion will now be derived. The inertia-gravity waves will be derived in both a non-hydrostatic and hydrostatic context. The scaling of non-hydrostatic effects is described in appendix B.

2.4.1 Non-hydrostatic Inertia-Gravity waves

Gravity waves are responsible for radiating energy away from regions of forcing and bringing the fluid into a geostrophically adjusted state. Rotation influences gravity waves producing dispersive inertia-gravity waves.

The modification to gravity waves by the rotation of the system is most easily demonstrated on an f-plane that assumes the planetary vorticity is constant in space and points along the local vertical. Ignoring the non-linear

terms in the momentum equations and assuming a representative background stratification $\frac{\partial \bar{\rho}}{\partial z} = \frac{-\rho_o}{g} N^2$, the single constituent, linearised system, in dimensional form is:

$$\nabla^2 p' + \rho_o f_o (\nabla \wedge \mathbf{u}) \cdot \hat{\mathbf{k}} + g \frac{\partial \rho'}{\partial z} = 0 \quad (2.36a)$$

$$\frac{\partial \mathbf{u}_h}{\partial t} + f_o \hat{\mathbf{k}} \wedge \mathbf{u}_h + \frac{1}{\rho_o} \nabla_h p' = 0 \quad (2.36b)$$

$$\nabla \cdot \mathbf{u} = 0 \quad (2.36c)$$

$$\frac{\partial \rho'}{\partial t} + w \frac{\partial \bar{\rho}}{\partial z} = 0 \quad (2.36d)$$

A succinct method for finding the natural modes of a system is to write the linear system as an amplifying matrix, in this instance:

$$\begin{pmatrix} \nabla^2 & \rho_o f_o \frac{\partial}{\partial y} & -\rho_o f_o \frac{\partial}{\partial x} & 0 & g \frac{\partial}{\partial z} \\ \frac{1}{\rho_o} \frac{\partial}{\partial x} & \frac{\partial}{\partial t} & -f_o & 0 & 0 \\ \frac{1}{\rho_o} \frac{\partial}{\partial y} & f_o & \frac{\partial}{\partial t} & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} & 0 \\ 0 & 0 & 0 & \frac{-\rho_o N^2}{g} & \frac{\partial}{\partial t} \end{pmatrix} \begin{pmatrix} p' \\ u \\ v \\ w \\ \rho' \end{pmatrix} = \mathbf{0} \quad (2.37)$$

Assuming a local solution of the form:

$$\begin{pmatrix} p' \\ u \\ v \\ w \\ \rho' \end{pmatrix} = \begin{pmatrix} p'_o \\ u_o \\ v_o \\ w_o \\ \rho'_o \end{pmatrix} e^{i(kx+ly+mz-\omega t)} \quad (2.38)$$

then on substitution into the linear system:

$$\det \begin{pmatrix} -(k^2 + l^2 + m^2) & i\rho_o f_o l & -i\rho_o f_o k & 0 & igm \\ \frac{i}{\rho_o} k & -i\omega & -f_o & 0 & 0 \\ \frac{i}{\rho_o} l & f_o & -i\omega & 0 & 0 \\ 0 & ik & il & im & 0 \\ 0 & 0 & 0 & \frac{-\rho_o N^2}{g} & -i\omega \end{pmatrix} = 0 \quad (2.39)$$

or

$$\omega m ((k^2 + l^2 + m^2)\omega^2 - m^2 f^2 - (k^2 + l^2)N^2) = 0 \quad (2.40)$$

There are three roots corresponding to the three natural modes. The trivial root $\omega = 0$ reflects the steady state of the geostrophic mode (steady because of the f-plane assumption). The remaining pair of roots give the dispersion relation for the non-hydrostatic gravity waves:

$$\omega^2 = \frac{m^2 f^2 + (k^2 + l^2)N^2}{k^2 + l^2 + m^2} \quad (2.41)$$

In the long horizontal wave limit, where the vertical scales of the wave motion are assumed much shorter than the lateral scales (a consequence of the aspect ratio of the ocean), then the vertical wave number will be much larger than the horizontal wave numbers, $\frac{k^2 + l^2}{m^2} \ll 1$. Re-writing the dispersion relation for non-hydrostatic inertia-gravity waves, an approximate form for the horizontally long waves is readily found:

$$\omega^2 = \frac{f^2 + \frac{k^2 + l^2}{m^2} N^2}{1 + \frac{k^2 + l^2}{m^2}} \simeq f^2 + (k^2 + l^2) \frac{N^2}{m^2} \quad (2.42)$$

The later form, applicable to the horizontally long waves, is similar to the form of the dispersion of the external inertia-gravity wave modes (interpreting N/m as \sqrt{gH} the gravity wave speed). This means that lessons learned about the numerical integration of the shallow water system (to be described in a later chapter) should be relevant to the treatment of internal inertia-gravity waves in more comprehensive models.

2.4.2 Hydrostatic approximation: Hydrostatic Inertia-Gravity waves

As just discussed, for motion with a small aspect ratio or indeed for nearly any motion in a stratified fluid, the horizontal gradient of non-hydrostatic pressure, vertical acceleration and horizontal coriolis terms can be neglected. The vertical momentum equation is reduced to the hydrostatic balance equation:

$$g\rho' + \frac{\partial p'}{\partial z} = 0 \quad (2.43)$$

Following the linearization procedure as for the non-hydrostatic gravity waves, the system can be approximated:

$$\frac{\partial^2 p'}{\partial z^2} + g\frac{\partial \rho'}{\partial z} = 0 \quad (2.44a)$$

$$\frac{\partial \mathbf{u}_h}{\partial t} + f_o \hat{\mathbf{k}} \wedge \mathbf{u}_h + \frac{1}{\rho_o} \nabla_h p' = 0 \quad (2.44b)$$

$$\nabla \cdot \mathbf{u} = 0 \quad (2.44c)$$

$$\frac{\partial \rho'}{\partial t} + w \frac{\partial \bar{\rho}}{\partial z} = 0 \quad (2.44d)$$

Again, the amplifying matrix method can be used to derive the natural modes of motions:

$$\det \left(\begin{array}{ccccc} -m^2 & 0 & 0 & 0 & \imath gm \\ \frac{\imath}{\rho_o} k & -\imath \omega & -f_o & 0 & 0 \\ \frac{\imath}{\rho_o} l & f_o & -\imath \omega & 0 & 0 \\ 0 & \imath k & \imath l & \imath m & 0 \\ 0 & 0 & 0 & \frac{-\rho_o N^2}{g} & -\imath \omega \end{array} \right) = 0 \quad (2.45)$$

or

$$\omega m (m^2(\omega^2 - f^2) - (k^2 + l^2)N^2) = 0 \quad (2.46)$$

Again, three roots exists, one pertaining to the steady geostrophic state. The remaining pair of roots correspond to hydrostatic inertia-gravity waves:

$$\omega^2 = f^2 + (k^2 + l^2) \frac{N^2}{m^2} \quad (2.47)$$

which is a slightly simpler form than the non-hydrostatic inertia-gravity waves. Note, however, that the approximated dispersion relation for horizontally long waves of the non-hydrostatic model corresponds to that of the hydrostatic gravity waves. This is self-consistent in that the hydrostatic approximation was made in the limit of small aspect ratio of motion.

2.4.3 Rossby Waves

Rossby waves are motions deriving from the slow evolution of the geostrophically adjusted fluid. The conventional derivation is given in the context of the shallow water equations in a later chapter. Here the more general method described above will be used to derive all three natural modes together; the pair of inertia-gravity waves and the Rossby wave.

Ignoring non-linearities, the equations describing the evolution of horizontal divergence, $D = \nabla_h \cdot \mathbf{u}_h$, and vorticity, $\zeta = \hat{\mathbf{k}} \cdot (\nabla_h \wedge \mathbf{u}_h)$, are:

$$\frac{\partial D}{\partial t} - f\zeta + \beta u + \frac{1}{\rho_o} \nabla_h^2 p = 0 \quad (2.48)$$

$$\frac{\partial \zeta}{\partial t} + fD + \beta v = 0 \quad (2.49)$$

Replacing the momentum equations with the above divergence and vorticity equations, and adding the definitions

of divergence and vorticity, the system expressed in the amplifying matrix form is:

$$\begin{pmatrix} -m^2 & 0 & 0 & 0 & \imath gm & 0 & 0 \\ \frac{-1}{\rho_0} |\hat{\mathbf{k}}|^2 & -\imath\omega & -f & 0 & 0 & \beta & 0 \\ 0 & f & -\imath\omega & 0 & 0 & 0 & \beta \\ 0 & 1 & 0 & \imath m & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-\rho_0 N^2}{g} & -\imath\omega & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & \imath k & \imath l \\ 0 & 0 & -1 & 0 & 0 & -\imath l & \imath k \end{pmatrix} \begin{pmatrix} p' \\ D \\ \zeta \\ w \\ \rho' \\ u \\ v \end{pmatrix} = 0 \quad (2.50)$$

The determinant of which gives:

$$\frac{\omega^3}{f^3} + \frac{2\beta k}{f|\mathbf{k}|^2} \frac{\omega^2}{f^2} - \left(1 + L_\rho |\mathbf{k}|^2 - \imath \frac{\beta l}{f|\mathbf{k}|^2}\right) \frac{\omega}{f} - \frac{\beta L_\rho}{f} L_\rho k = 0 \quad (2.51)$$

where $L_\rho^2 = N^2/(f^2 m^2)$ is the square of the Rossby radius of deformation.

In the special case of β vanishing, the steady root and a pair of hydrostatic inertia-gravity waves are recovered. More generally, for motion such that $\frac{\beta}{f|\mathbf{k}|} \ll 1$ (*i.e.* the wave lengths are small compared to the planetary radius, the distance over which planetary vorticity varies) then the inertia-gravity modes are the approximate non-zero roots of the above dispersion relation:

$$\frac{\omega^2}{f^2} \approx 1 + L_\rho^2 (k^2 + l^2) \quad (2.52)$$

The near zero root, that has appeared consistently as a zero root to this point, can be obtained by assuming that the frequency is small compared to the inertial frequency (this is consistent with it being a slow motion with a near zero frequency). Neglecting terms of second order or higher leaves the dispersion relation:

$$\omega = \frac{-\beta L_\rho^2 k}{1 + L_\rho^2 |\mathbf{k}|^2 - \imath \frac{\beta l}{f|\mathbf{k}|^2}} \quad (2.53)$$

Again, if the wave lengths are shorter than the planetary scale, the dispersion relation simplifies to:

$$\omega \approx \frac{-\beta k}{\frac{1}{L_\rho^2} + |\mathbf{k}|^2} \quad (2.54)$$

It should be clear that for the long waves that do not satisfy these assumptions exactly, there is a significant complex component to the frequency that indicates an exponential type of behaviour. The simple dispersion relations derived conventionally are thus only approximate and most accurate for the shorter waves. Nevertheless, a great deal of the ocean circulation can be understood in terms of the plain wave propagation described here.

More importantly for this study, the linear wave-like behaviour of these natural modes of motion should be reproduced by numerical models that claim to explicitly resolve the processes involved. As will be seen, models often fail to accurately represent these motions.

2.5 Summary and comments

The Navier-Stokes equations, that describe all classes of motion in the ocean, were derived and later summarised in section 2.1. Ill conditioning and subsequent scaling of the equations justified making the Boussinesq approximation. The fast acoustic modes were filtered by making the flow non-divergent (incompressible). The non-hydrostatic version of the model is then Boussinesq and incompressible.

Except where the stratification and aspect ratio of the flow are sufficiently small, the hydrostatic approximation is quite valid. Inertia-gravity waves are superficially modified by the hydrostatic approximation if the horizontal wave length is sufficiently short.

The Rossby wave and pair of inertia-gravity waves make up the three natural modes of the single constituent fluid (*i.e.* the two thermodynamic variables, potential temperature and salinity, are replaced with one variable, potential density). It was stated that these waves motions are responsible for setting up many of the features present in the circulation of the world oceans. For instance, the anisotropic propagation of Rossby waves gives rise to the Western intensification of boundary currents.

Most important to this study, the proper behaviour of these natural modes of motion should be reproduced by numerical models that claim to explicitly resolve the processes involved. As will be seen, models often fail to accurately represent these motions. In particular, the shortest resolvable inertia-gravity waves in the model to be described in chapter 3, fail to oscillate inertially. This makes the model particularly susceptible to grid-scale noise. Much of this study is devoted to developing a solution to this problem.

Chapter 3

A Navier-Stokes Ocean Model

Here, the numerical model based on the non-hydrostatic, incompressible Navier-Stokes equations (2.11) is described. First, the terminology which is used throughout this and all subsequent chapters is established. Then the continuous and discrete formulations of the model. The conservation properties of the discrete model are described. The model is then applied and the strengths and weaknesses of it reviewed. Examples of applications and the problematic results associated with the Coriolis term are shown.

Chapters 4 and 5 discuss the problems associated with the Coriolis term on a C grid and present a correction to this problem. Chapter 6 discusses the re-formulation of the model using finite volumes, and the ability to shave cells to represent topography.

3.1 Finite difference methods

Finite differencing tries to reduce the truncation error in the evaluation of the governing equations. To express the equations in terms of the discrete dependent variables, the variables are linked through Taylor expansions about appropriate points in space (or time if the method is being applied to the time-stepping).

For example, let the discrete dependent variables f_i be carried at discretely separated positions x_i , with intervals $\Delta x_{i+\frac{1}{2}}$ (see figure 3.1). The discrete variables are assumed to match the continuous function that they represent at the appropriate points.

The Taylor expansion of the continuous function about x_i is:

$$f_{i\pm 1} = f_i \pm f'(x_i)\Delta x_{i\pm\frac{1}{2}} + f''(x_i)\frac{\Delta x_{i\pm\frac{1}{2}}^2}{2!} \pm f'''(x_i)\frac{\Delta x_{i\pm\frac{1}{2}}^3}{3!} + \dots \quad (3.1)$$

The spatial derivative, at x_i , of the continuous function, $f'(x) = \frac{\partial f}{\partial x}$, can be approximated in terms of the discrete variables f_{i-1} , f_i and f_{i+1} :

$$f'(x_i) = \frac{f_{i+1} - f_i}{\Delta x_{i+\frac{1}{2}}} - \frac{\Delta x_{i+\frac{1}{2}} f''(x_i)}{2!} - \frac{\Delta x_{i+\frac{1}{2}}^2 f'''(x_i)}{3!} - \dots \quad (3.2)$$

or

$$f'(x_i) = \frac{f_i - f_{i-1}}{\Delta x_{i-\frac{1}{2}}} + \frac{\Delta x_{i-\frac{1}{2}} f''(x_i)}{2!} - \frac{\Delta x_{i-\frac{1}{2}}^2 f'''(x_i)}{3!} + \dots \quad (3.3)$$

Truncating this expression to the known quantities, *i.e.* neglecting terms involving higher derivatives of f , yields what is often called ‘side differencing’:

$$f'(x_i) \approx \frac{f_{i+1} - f_i}{\Delta x_{i+\frac{1}{2}}} \quad \text{or} \quad f'(x_i) \approx \frac{f_i - f_{i-1}}{\Delta x_{i-\frac{1}{2}}} \quad (3.4)$$

where the truncation errors are of order $O(\frac{1}{2}\Delta x)$. This is termed first order accurate, referring to the power of Δx in the truncation error. As the resolution of the model is increased, the truncation error gets smaller and the discrete model approaches the continuous system.

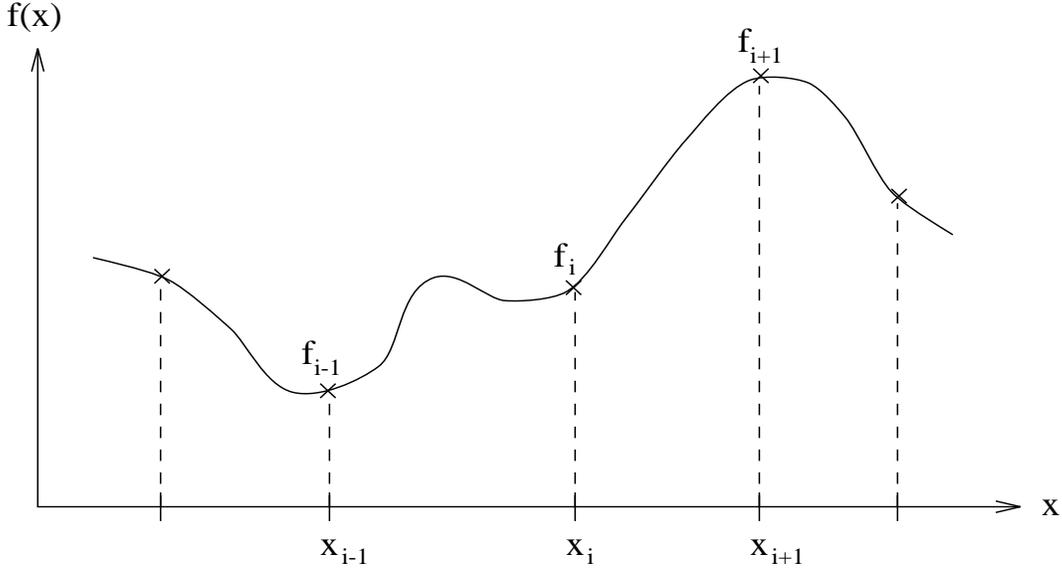


Figure 3.1: The continuous function $f(x)$ is described by the discrete variables f_i that match the function at the discrete position x_i .

Returning to the two untruncated Taylor expansions, they can be combined to eliminate the second derivative terms leaving second order truncation terms:

$$f'(x_i) \approx \frac{\frac{\Delta x_{i-\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}} f_{i+1} + \left(\frac{\Delta x_{i+\frac{1}{2}}}{\Delta x_{i-\frac{1}{2}}} - \frac{\Delta x_{i-\frac{1}{2}}}{\Delta x_{i+\frac{1}{2}}} \right) f_i - \frac{\Delta x_{i+\frac{1}{2}}}{\Delta x_{i-\frac{1}{2}}} f_{i-1}}{\Delta x_{i+\frac{1}{2}} + \Delta x_{i-\frac{1}{2}}} \quad (3.5)$$

Here the truncation error is $O(\frac{1}{3!} \Delta x_{i-\frac{1}{2}} \Delta x_{i+\frac{1}{2}})$. For the less general case of regular grid spacing, $\Delta x = \Delta x_{i-\frac{1}{2}} = \Delta x_{i+\frac{1}{2}}$, then the scheme reduces to a more intuitive form:

$$f'(x_i) \approx \frac{f_{i+1} - f_{i-1}}{2\Delta x} \quad (3.6)$$

This is referred to as ‘centered differencing’ and is clearly preferable to ‘side differencing’ due to the dramatic decrease in truncation error; $O(\frac{1}{2}\Delta x) \gg O(\frac{1}{3!}\Delta x^2)$. To obtain an equivalent accuracy with the first order scheme, as the second order scheme with N points, one would require $3N^2$ points. The dramatic improvement in accuracy is due to the centered evaluation of the gradient.

A further improvement in accuracy of a model can be obtained by ‘staggering’ model variables. In general, odd powered derivatives are staggered with the even powers. For example, consider evaluating $\frac{\partial f}{\partial x}$ in terms of f_i and f_{i+1} at a position $x = \alpha x_i + (1 - \alpha)x_{i+1}$, where $0 \leq \alpha \leq 1$ which lies on or between the two nodes. Taylor expansion about x and eliminating the undefined $f(x)$ terms yields:

$$\frac{\partial f}{\partial x} \Big|_{\alpha} = \frac{f_{i+1} - f_i}{\Delta x} - \frac{(1 - 2\alpha)\Delta x}{2!} f'' - \frac{(1 - 3\alpha + 3\alpha^2)\Delta x^2}{3!} f''' - \dots \quad (3.7)$$

The first truncation term indicates first order accuracy. The limits of $\alpha = 0, 1$ correspond to the first order accurate side differencing described earlier. The special case of $\alpha = \frac{1}{2}$, where the staggering is centered, causes the first truncation term to vanish where upon the scheme becomes second order accurate with a factor of four improvement over the previous second order scheme; $O(\frac{1}{4 \cdot 3!} \Delta x^2)$:

$$f'(x_{i+\frac{1}{2}}) \approx \frac{f_{i+1} - f_i}{\Delta x} \quad (3.8)$$

Staggered second order accurate finite differencing involves the same number of points in a finite difference stencil as first order side differencing, and less points than centered unstaggered differencing. It is the most accurate second order scheme and uses the smallest stencil.

3.2 Finite difference notation and rules

Before describing the spatial discretisation of the model, some notation and elementary relations will be established. The notation used here is based upon that used by Arakawa and Lamb [AL77].

The centered, staggered finite difference operators will be denoted:

$$\delta_x \phi \equiv \phi_{i+\frac{1}{2},j,k} - \phi_{i-\frac{1}{2},j,k} \quad (3.9a)$$

$$\delta_y \phi \equiv \phi_{i,j+\frac{1}{2},k} - \phi_{i,j-\frac{1}{2},k} \quad (3.9b)$$

$$\delta_z \phi \equiv \phi_{i,j,k+\frac{1}{2}} - \phi_{i,j,k-\frac{1}{2}} \quad (3.9c)$$

and the respective interpolation or averaging operators:

$$\overline{\phi}^x \equiv \frac{\phi_{i+\frac{1}{2},j,k} + \phi_{i-\frac{1}{2},j,k}}{2} \quad (3.10a)$$

$$\overline{\phi}^y \equiv \frac{\phi_{i,j+\frac{1}{2},k} + \phi_{i,j-\frac{1}{2},k}}{2} \quad (3.10b)$$

$$\overline{\phi}^z \equiv \frac{\phi_{i,j,k+\frac{1}{2}} + \phi_{i,j,k-\frac{1}{2}}}{2} \quad (3.10c)$$

Staggered, second order differencing for the node i will be written:

$$\frac{\phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}}{\Delta x_i} = \frac{1}{\Delta x} \delta_x \phi \quad (3.11)$$

where Δx is defined for the interval i .

The difference and interpolation operators can be shown to satisfy the following rules:

$$\delta_\zeta \delta_\eta \phi = \delta_\eta \delta_\zeta \phi \quad (3.12a)$$

$$\delta_\zeta \overline{\phi}^\eta = \overline{\delta_\zeta \phi}^\eta \quad (3.12b)$$

$$\overline{\overline{\phi}}^\zeta = \overline{\phi}^\zeta \quad (3.12c)$$

$$\delta_\zeta(\phi\psi) = \overline{\phi}^\zeta \delta_\zeta \psi + \overline{\psi}^\zeta \delta_\zeta \phi \quad (3.12d)$$

$$\delta_\zeta(\overline{\phi}^\zeta \psi) = \phi \delta_\zeta \psi + \overline{\psi} \delta_\zeta \overline{\phi}^\zeta \quad (3.12e)$$

$$\overline{\phi\psi}^\zeta = \overline{\phi}^\zeta \overline{\psi}^\zeta + \frac{1}{4} \delta_\zeta \phi \delta_\zeta \psi \quad (3.12f)$$

$$\overline{\overline{\phi}^\zeta \psi}^\zeta = \overline{\phi\psi}^\zeta + \frac{1}{4} \delta_\zeta(\psi \delta_\zeta \phi) \quad (3.12g)$$

where ζ and η can be any coordinate and need not be different. ϕ and ψ are model variable or expressions that *must* be evaluated at the same points in the model.

One further piece of short-hand that is not conventional is:

$$\wr \phi \wr_x^2 \equiv \phi_{i-\frac{1}{2},j,k} \phi_{i+\frac{1}{2},j,k} \quad (3.13a)$$

$$\wr \phi \wr_y^2 \equiv \phi_{i,j-\frac{1}{2},k} \phi_{i,j+\frac{1}{2},k} \quad (3.13b)$$

$$\wr \phi \wr_z^2 \equiv \phi_{i,j,k-\frac{1}{2}} \phi_{i,j,k+\frac{1}{2}} \quad (3.13c)$$

which satisfies the relation:

$$\overline{\overline{\phi}^\zeta \overline{\phi}^\zeta} = \frac{1}{2} \overline{\phi^2}^\zeta + \frac{1}{2} \wr \phi \wr_\zeta^2 \quad (3.14)$$

This last product operator is introduced to keep the notation concise when conservation of second moments is derived later. The operator $\sqrt{\frac{1}{2} \wr \phi \wr_\zeta^2}$ is the geometric mean between two neighbouring points.

3.3 Continuous formulation of model

The Inviscid, Adiabatic and Incompressible Boussinesq equations of motion were derived in chapter 2. The process of discretisation introduces sub-grid scale eddy terms that have to be parameterised in order to close the system. The

non-divergence of the flow allows the Lagrangian advective operator to be written as the divergence of an advective flux. The continuous equations become:

$$\frac{\partial u}{\partial t} + \frac{1}{\rho_o} \frac{\partial}{\partial x} (p_s + p_{nh}) = G_u \quad (3.15a)$$

$$\frac{\partial v}{\partial t} + \frac{1}{\rho_o} \frac{\partial}{\partial y} (p_s + p_{nh}) = G_v \quad (3.15b)$$

$$\frac{\partial w}{\partial t} + \frac{1}{\rho_o} \frac{\partial}{\partial z} (p_s + p_{nh}) = G_w \quad (3.15c)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \quad (3.15d)$$

$$\frac{\partial \theta}{\partial t} = G_\theta \quad (3.15e)$$

$$\frac{\partial S}{\partial t} = G_s \quad (3.15f)$$

$$\rho' = \rho(\theta, S, -g\rho_o z) - \rho_o \quad (3.15g)$$

$$\frac{\partial}{\partial z} p_h = -g\rho' \quad (3.15h)$$

where the source terms or G s are given by:

$$G_u = +2\Omega(v \sin \phi - w \cos \phi) - \nabla \cdot (u\mathbf{u}) - \frac{1}{\rho_o} \frac{\partial}{\partial x} p_h + \frac{1}{\rho_o} \frac{\partial \tau^{(x)}}{\partial z} + \nabla \cdot (\boldsymbol{\nu} \nabla u) \quad (3.16a)$$

$$G_v = -2\Omega u \sin \phi - \nabla \cdot (v\mathbf{u}) - \frac{1}{\rho_o} \frac{\partial}{\partial y} p_h + \frac{1}{\rho_o} \frac{\partial \tau^{(y)}}{\partial z} + \nabla \cdot (\boldsymbol{\nu} \nabla v) \quad (3.16b)$$

$$G_w = +2\Omega u \cos \phi - \nabla \cdot (w\mathbf{u}) - \nabla \cdot (\boldsymbol{\nu} \nabla w) \quad (3.16c)$$

$$G_\theta = -\nabla \cdot (\theta\mathbf{u} - \boldsymbol{\kappa}_\theta \nabla \theta) + \mathcal{H} \quad (3.16d)$$

$$G_s = -\nabla \cdot (S\mathbf{u} - \boldsymbol{\kappa}_s \nabla S) + \mathcal{Q}_s \quad (3.16e)$$

$$(3.16f)$$

The rigid lid approximation and no flux through solid boundaries is expressed:

$$w(z=0) = \mathbf{u} \cdot \hat{\mathbf{n}} = 0 \quad (3.17)$$

Making use of the continuity equation, diagnostic equations for both the surface pressure and non-hydrostatic pressure can be derived from the momentum equations. The surface pressure equation is only accurate in the hydrostatic limit of the model:

$$\nabla_h \cdot H \nabla_h p_s = \frac{\partial}{\partial x} \int_{-H(x,y)}^0 G_u dz + \frac{\partial}{\partial y} \int_{-H(x,y)}^0 G_v dz \quad (3.18)$$

For the non-hydrostatic model, a further elliptic equation for the non-hydrostatic pressure is solved. Included in the source term of the equation is the residual from the 2-D inversion to correct for any inaccuracies:

$$\nabla^2 p_{nh} = \frac{\partial}{\partial x} G_u + \frac{\partial}{\partial y} G_v + \frac{\partial}{\partial z} G_w - \left(\nabla_h^2 p_s - \nabla_h \cdot \overline{\mathbf{G}_h^z} \right) \quad (3.19)$$

3.4 Spatial discretisation of model

The finite differencing will be described in two sections; one describing the spatial distribution and finite difference schemes, the second will be concerned with the time-stepping and related issues.

The model variables are staggered in the three dimensional equivalent of an Arakawa C grid (see figure 3.2). The tracers are all carried at the p point.

Notice that the staggering of variables both introduces and removes the need for spatial interpolation of different terms in the model. For example, the advective flux of tracers need not involve interpolation of the flow whilst evaluation of the Coriolis terms involves spatial averaging.

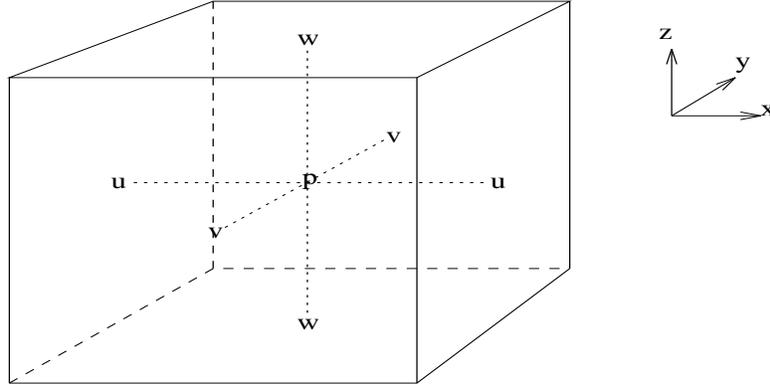


Figure 3.2: The three dimensional distribution of model variables. The variables S and θ are also carried at the p points.

The discrete hydrostatic equation is:

$$\delta_z p_h = -\overline{g\rho'\Delta z^z} \quad (3.20)$$

and can be integrated from the top down using the upper boundary condition of $p_h(0) = p_{Atmos}$.

The discretisation of the momentum equations is:

$$\frac{\partial u}{\partial t} = \frac{-1}{\rho_o \Delta x} \delta_x (p_s + p_{nh}) + G_u \quad (3.21a)$$

$$\frac{\partial v}{\partial t} = \frac{-1}{\rho_o \Delta y} \delta_y (p_s + p_{nh}) + G_v \quad (3.21b)$$

$$\frac{\partial w}{\partial t} = \frac{-1}{\rho_o \Delta z} \delta_z (p_s + p_{nh}) + G_w \quad (3.21c)$$

where the G s are discretised as:

$$\begin{aligned} G_u &= 2\Omega \left(\begin{array}{l} \overline{v^{xy}} \sin \phi \\ + \overline{w^{xz}} \cos \phi \end{array} \right) \begin{array}{l} -\frac{1}{\Delta x} \delta_x (\overline{u^x \overline{u^x}}) \\ -\frac{1}{\Delta y} \delta_y (\overline{v^y \overline{u^y}}) \\ -\frac{1}{\Delta z} \delta_z (\overline{w^z \overline{u^z}}) \end{array} - \frac{1}{\rho_o \Delta x} \delta_x p_h + \frac{1}{\rho_o \Delta z} \delta_z \tau^{(x)} + \nabla \cdot \nu \nabla u \\ G_v &= -2\Omega \overline{v^x} \sin \phi^y \begin{array}{l} -\frac{1}{\Delta x} \delta_x (\overline{u^y \overline{v^x}}) \\ -\frac{1}{\Delta y} \delta_y (\overline{v^y \overline{v^y}}) \\ -\frac{1}{\Delta z} \delta_z (\overline{w^y \overline{v^z}}) \end{array} - \frac{1}{\rho_o \Delta y} \delta_y p_h + \frac{1}{\rho_o \Delta z} \delta_z \tau^{(y)} + \nabla \cdot \nu \nabla v \\ G_w &= +2\Omega \overline{u^{xy}} \cos \phi \begin{array}{l} -\frac{1}{\Delta x} \delta_x (\overline{u^z \overline{w^x}}) \\ -\frac{1}{\Delta y} \delta_y (\overline{v^z \overline{w^y}}) \\ -\frac{1}{\Delta z} \delta_z (\overline{w^z \overline{w^z}}) \end{array} + \nabla \cdot \nu \nabla w \end{aligned} \quad (3.22)$$

and ϕ is evaluated at p latitudes. The eddy viscosity terms have the form:

$$\nabla \cdot \nu \nabla \phi = \frac{1}{\Delta x^2} \delta_x \nu^{(x)} \delta_x \phi + \frac{1}{\Delta y^2} \delta_y \nu^{(y)} \delta_y \phi + \frac{1}{\Delta z^2} \delta_z \nu^{(z)} \delta_z \phi$$

but they can be made more general.

The contributions to the rate of change of zonal kinetic energy is obtained by multiplying G_u by u . The vertical component of Coriolis introduces a portion from each of four neighbouring v points:

$$\frac{1}{4} 2\Omega \sin \phi_j v_{i\pm 1, j \pm \frac{1}{2}, k} u_{i + \frac{1}{2}, j, k} \quad (3.23)$$

The contribution of the meridional velocity to the kinetic energy at one of the four neighbouring v points, $i, j + \frac{1}{2}$, is:

$$\frac{-1}{4} 2\Omega \sin \phi_j u_{i + \frac{1}{2}, j, k} v_{i, j + \frac{1}{2}, k} \quad (3.24)$$

In the summation over the entire model, the vertical Coriolis terms as written do not violate the kinetic energy budget. A similar analysis shows that this is true of the horizontal Coriolis terms and pressure gradient terms also. The eddy viscosity terms do extract energy from the system, as required.

The depth integrated flow is non-divergent (since the model has a rigid lid):

$$\frac{1}{\Delta x} \delta_x \int_{H_u}^0 u \, dz + \frac{1}{\Delta y} \delta_y \int_{H_v}^0 v \, dz = 0 \quad (3.25)$$

where the boundary conditions are implicitly satisfied wherever $H = 0$.

Substituting in the discrete momentum equations gives:

$$\frac{1}{\Delta x} \delta_x H_u \frac{1}{\Delta x} \delta_x p_s + \frac{1}{\Delta y} \delta_y H_v \frac{1}{\Delta y} \delta_y p_s = \frac{1}{\Delta x} \delta_x \int_{H_u}^0 G_u \, dz + \frac{1}{\Delta y} \delta_y \int_{H_v}^0 G_v \, dz \quad (3.26)$$

which has ignored any accelerations due to non-hydrostatic pressure gradients.

The local non-divergence of the flow is represented discretely as:

$$\frac{1}{\Delta x} \delta_x u + \frac{1}{\Delta y} \delta_y v + \frac{1}{\Delta z} \delta_z w = 0 \quad (3.27)$$

and the boundary conditions of no normal flow at solid boundaries is applied here.

Again, substituting in the discrete momentum equations gives:

$$\begin{aligned} & \left(\frac{1}{\Delta x^2} \delta_x \delta_x p_{nh} + \frac{1}{\Delta y^2} \delta_y \delta_y p_{nh} + \frac{1}{\Delta z^2} \delta_z \delta_z p_{nh} \right) = \\ & + \left(\frac{1}{\Delta x} \delta_x (G_u - \frac{1}{\Delta x} \delta_x p_s) + \frac{1}{\Delta y} \delta_y (G_v - \frac{1}{\Delta y} \delta_y p_s) + \frac{1}{\Delta z} \delta_z G_w \right) \end{aligned} \quad (3.28)$$

This last equation is only solved if the model is in the non-hydrostatic regime.

The discretisation of the thermodynamic G_θ and G_s are:

$$G_\theta = \begin{aligned} & -\frac{1}{\Delta x^2} \delta_x (u \bar{\theta}^x - \kappa_\theta^{(x)} \delta_x \theta) \\ & -\frac{1}{\Delta y^2} \delta_y (u \bar{\theta}^y - \kappa_\theta^{(y)} \delta_y \theta) + \mathcal{H} \\ & -\frac{1}{\Delta z^2} \delta_z (u \bar{\theta}^z - \kappa_\theta^{(z)} \delta_z \theta) \end{aligned} \quad (3.29a)$$

$$G_s = \begin{aligned} & -\frac{1}{\Delta x^2} \delta_x (u \bar{S}^x - \kappa_s^{(x)} \delta_x S) \\ & -\frac{1}{\Delta y^2} \delta_y (u \bar{S}^y - \kappa_s^{(y)} \delta_y S) + \mathcal{Q}_s \\ & -\frac{1}{\Delta z^2} \delta_z (u \bar{S}^z - \kappa_s^{(z)} \delta_z S) \end{aligned} \quad (3.29b)$$

3.4.1 Conservative advection

For non-divergent flow, the continuous equations have the property that all moments of a conserved tracer, ϕ , are conserved:

$$\left. \begin{aligned} \frac{\partial \phi}{\partial t} + \nabla \cdot \phi \mathbf{u} &= 0 \\ \nabla \cdot \mathbf{u} &= 0 \end{aligned} \right\} \Rightarrow \frac{\partial \phi^n}{\partial t} + \nabla \cdot \phi^n \mathbf{u} = 0 \quad \forall n = 1, 2, \dots \quad (3.30)$$

The advection operator is written as the divergence of advective fluxes

$$\frac{\partial \phi}{\partial t} + \frac{1}{\Delta x} \delta_x (\bar{\phi}^x u) + \frac{1}{\Delta y} \delta_y (\bar{\phi}^y v) + \frac{1}{\Delta z} \delta_z (\bar{\phi}^z w) = \dots \quad (3.31)$$

which guarantees conservation of the total first moment of the scalar in the discrete model:

$$\frac{\partial}{\partial t} \iiint \phi \, dV = \frac{\partial}{\partial t} \sum_{i,j,k} \phi_{i,j,k} = 0 \quad (3.32)$$

The x component of the flux form can be re-arranged using rule 3.12e:

$$\delta_x(\overline{\phi^x u}) = \phi \delta_x u + \overline{u \delta_x \phi^x} \quad (3.33)$$

and the same can be done for the y and z terms. Substituting into the advection term gives:

$$\frac{\partial \phi}{\partial t} + \frac{1}{\Delta x} \overline{u \delta_x \phi^x} + \frac{1}{\Delta y} \overline{v \delta_y \phi^y} + \frac{1}{\Delta z} \overline{w \delta_z \phi^z} + \phi \left(\frac{\delta_x u}{\Delta x} + \frac{\delta_y v}{\Delta y} + \frac{\delta_z w}{\Delta z} \right) = \dots \quad (3.34)$$

where the last term vanishes if the flow is non-divergent according to this definition.

Multiplying by ϕ gives:

$$\frac{\partial \phi^2}{\partial t} + \frac{1}{\Delta x} \overline{\phi u \delta_x \phi^x} + \frac{1}{\Delta y} \overline{\phi v \delta_y \phi^y} + \frac{1}{\Delta z} \overline{\phi w \delta_z \phi^z} = \dots \quad (3.35)$$

Considering only the x direction, the advective term can be written:

$$\overline{\phi u \delta_x \phi^x} = \overline{u \overline{\phi^x} \delta_x \phi^x} - \frac{1}{4} \delta_x (u \delta_x \phi \delta_x \phi) \quad (3.36)$$

where use has been made of 3.12g. The last term can be re-written using a special case of 3.12f:

$$-\frac{1}{4} \delta_x (u \delta_x \phi \delta_x \phi) = \delta_x \left(u (\overline{\phi^x \phi^x} - \overline{\phi^2}^x) \right) \quad (3.37)$$

and then introducing the unconventional notation via 3.14:

$$\overline{\phi u \delta_x \phi^x} = \delta_x \left(u \frac{1}{2} \wr \phi \wr_x^2 \right) - \frac{\phi^2}{2} \delta_x u \quad (3.38)$$

Repeating the procedure for the other two contributions gives a flux form for the second moment of the scalar:

$$\frac{\partial \phi^2}{\partial t} + \frac{1}{\Delta x} \delta_x (u \wr \phi \wr_x^2) + \frac{1}{\Delta y} \delta_y (v \wr \phi \wr_y^2) + \frac{1}{\Delta z} \delta_z (w \wr \phi \wr_z^2) - \frac{\phi^2}{2} \left(\frac{\delta_x u}{\Delta x} + \frac{\delta_y v}{\Delta y} + \frac{\delta_z w}{\Delta z} \right) = \dots \quad (3.39)$$

where once again, if the flow is non-divergent the last term vanishes.

The same advection scheme is applied to the momentum equations, though here, the advecting flow must be interpolated appropriately. To advect u in the conservative manner just described, the advecting u^* , v^* and w^* must be interpolated from u , v and w in such a way as to be non-divergent.

$$u^* = \overline{u^x} \quad v^* = \overline{v^x} \quad w^* = \overline{w^x} \quad (3.40)$$

Expanding the divergence of the advecting flow and making use of the commutativity between interpolation and difference operators yields:

$$\frac{\delta_x u^*}{\Delta x} + \frac{\delta_y v^*}{\Delta y} + \frac{\delta_z w^*}{\Delta z} = \frac{\delta_x \overline{u^x}}{\Delta x} + \frac{\delta_y \overline{v^x}}{\Delta y} + \frac{\delta_z \overline{w^x}}{\Delta z} = \overline{\frac{\delta_x u}{\Delta x} + \frac{\delta_y v}{\Delta y} + \frac{\delta_z w}{\Delta z}} = 0 \quad (3.41)$$

Therefore, the total of zonal, meridional and vertical velocity contributions to the kinetic energy are conserved by the advection operator.

Higher moments of momentum and scalars are not conserved. Higher order differencing and forms of advection could perhaps be used to achieve this. Here, the efficiency and simplicity of the numerical model was deemed more important than conservation of higher moments.

3.5 The time-stepping scheme (Adams-Bashforth II)

Consider the horizontal momentum equations in an abbreviated form:

$$\frac{\partial}{\partial t} \mathbf{u} = \mathbf{P}(\mathbf{p}) + \mathbf{G}(\mathbf{u}, \mathbf{x}, t) \quad (3.42)$$

where \mathbf{P} contains the horizontal pressure gradient and is therefore a function only of the pressure field. All other terms that appear in the horizontal momentum equations are grouped into \mathbf{G} which will then be a function of the velocities, space and time.

To arrive at a time-stepping scheme, integrate equation 3.42 over a finite period of time, Δt :

$$\mathbf{u}(t_o + \Delta t) - \mathbf{u}(t_o) = + \int_{t_o}^{t_o + \Delta t} \mathbf{P}(t) dt + \int_{t_o}^{t_o + \Delta t} \mathbf{G}(t) dt \quad (3.43)$$

Here after, time levels $t_o - \Delta t$, $t_o - \frac{1}{2}\Delta t$, t_o , $t_o + \frac{1}{2}\Delta t$ and $t_o + \Delta t$ will be denoted by the superscripts $n-1$, $n-\frac{1}{2}$, n , $n+\frac{1}{2}$ and $n+1$ respectively. Due to the left hand side of 3.43, the velocities are defined at integer time levels, *i.e.* $n-1$, n , etc. The function $\mathbf{G}(t)$ is thus most naturally evaluated at these time levels while the pressure variable is free to be staggered in time (ie. carried at $n-\frac{1}{2}$, $n+\frac{1}{2}$ etc). The latter is convenient since the integral over the period n to $n+1$ of some function can most conveniently be approximated as Δt times the value of the function at the mid-point, $n+\frac{1}{2}$.

The integral of the function $\mathbf{G}(t)$ involves evaluations at the end points of the period and so the most accurate approximation involving just two time levels is a trapezoidal scheme. Here the integral of a function is Δt time the mean of the function evaluated at the two end points n and $n+1$. This is a semi-implicit method meaning that some of the unknown variables at $n+1$ are on the right hand side. Therefore, the equations must be re-arranged and solved as a set of simultaneous equations.

The function \mathbf{G} typically involves spatial operators and non-linear terms. A semi-implicit evaluation of this integral then requires sophisticated methods to solve for future variables.

Explicit evaluation methods for the integral of \mathbf{G} would appear to be cheaper and less complicated. The Adams-Bashforth method (AB2) is a quasi-second order method that involves a third time-level:

$$\begin{aligned} \int_{n\Delta t}^{(n+1)\Delta t} f(t) dt &\approx \Delta t (f^{(n)} + (\frac{1}{2} + \epsilon)\{f^{(n)} - f^{(n-1)}\}) \\ &= \Delta t ((\frac{3}{2} + \epsilon)f^{(n)} - (\frac{1}{2} + \epsilon)f^{(n-1)}) \end{aligned} \quad (3.44)$$

which amounts to a linear extrapolation in time to a point, $n + \frac{1}{2} + \epsilon$ that is just into the future of the mid-point $n+\frac{1}{2}$.

The final form of equation 3.43 using the above methods reads:

$$\begin{aligned} \frac{1}{\Delta t} (\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}) &= \mathbf{P}^{(n+\frac{1}{2})} \\ &+ (\frac{3}{2} + \epsilon)\mathbf{G}^{(n)} - (\frac{1}{2} + \epsilon)\mathbf{G}^{(n-1)} \end{aligned} \quad (3.45)$$

The tracer equations are stepped forward in a similar fashion:

$$\frac{1}{\Delta t} (\phi^{(n+1)} - \phi^{(n)}) = (\frac{3}{2} + \epsilon)\mathcal{S}^{(n)} - (\frac{1}{2} + \epsilon)\mathcal{S}^{(n-1)} \quad (3.46)$$

where \mathcal{S} contains all other terms.

3.6 Model Algorithm

The integration procedure is as follows:

1. Evaluate the density perturbation

$$\rho' = \rho(\theta^{(n)}, S^{(n)}, -g\rho_o z)$$

2. Integrate the hydrostatic equation for p_h

$$p_h = -g \int_{z=0}^z \bar{\rho}'^z dz$$

3. Evaluate $G_u^{(n)}$, $G_v^{(n)}$ (and $G_w^{(n)}$ if non-hydrostatic) using variables from time level n and the hydrostatic pressure.

4. Evaluate the accelerations

$$G_u^* = (\frac{3}{2} + \epsilon)G_u^{(n)} - (\frac{1}{2} + \epsilon)G_u^{(n-1)}$$

$$G_v^* = (\frac{3}{2} + \epsilon)G_v^{(n)} - (\frac{1}{2} + \epsilon)G_v^{(n-1)}$$

If non-hydrostatic

$$G_w^* = \left(\frac{3}{2} + \epsilon\right)G_w^{(n)} - \left(\frac{1}{2} + \epsilon\right)G_w^{(n-1)}$$

5. Solve the elliptic problem

$$\delta_x H_u \delta_x p_s + \delta_y H_v \delta_y p_s = \delta_x \int_{H_u}^0 \left(\frac{u^{(n)}}{\Delta t} + G_u^*\right) dz + \delta_y \int_{H_v}^0 \left(\frac{v^{(n)}}{\Delta t} + G_v^*\right) dz$$

for the surface pressure, p_s . This guarantees that the depth integrated flow will be non-divergent at the next time level.

6. If non-hydrostatic, solve the elliptic problem

$$[\delta_x \delta_x + \delta_y \delta_y + \delta_z \delta_z] p_{nh} = \delta_x \left(\frac{u^{(n)}}{\Delta t} + G_u^* - \delta_x p_s\right) + \delta_y \left(\frac{v^{(n)}}{\Delta t} + G_v^* - \delta_y p_s\right) + \delta_z \left(\frac{w^{(n)}}{\Delta t} + G_w^*\right)$$

for the non-hydrostatic pressure, otherwise let $p_{nh} = 0$.

7. Update horizontal velocities:

$$u^{(n+1)} = u^{(n)} + \Delta t \left(G_u^* - \frac{1}{\rho_o} \delta_x (p_s + p_{nh}) \right)$$

$$v^{(n+1)} = v^{(n)} + \Delta t \left(G_v^* - \frac{1}{\rho_o} \delta_y (p_s + p_{nh}) \right)$$

8. Integrate the continuity equation for w

$$w^{(n+1)} = - \int_H^z (\delta_x u^{(n+1)} + \delta_y v^{(n+1)}) dz$$

9. Evaluate the G s for θ and S using the latest velocities.

10. Update the thermodynamic variables

$$\theta^{(n+1)} = \theta^{(n)} + \Delta t \left(\left(\frac{3}{2} + \epsilon\right)G_\theta^{(n)} - \left(\frac{1}{2} + \epsilon\right)G_\theta^{(n-1)} + \mathcal{H}^{(\cdot)} \right)$$

$$S^{(n+1)} = S^{(n)} + \Delta t \left(\left(\frac{3}{2} + \epsilon\right)G_S^{(n)} - \left(\frac{1}{2} + \epsilon\right)G_S^{(n-1)} + \mathcal{Q}_S^{(\cdot)} \right)$$

11. Update indices: $n \rightarrow n + 1$ and repeat procedure for the next time-step.

3.7 3-D Elliptic Inversion

The diagnostic equation for pressure is a three dimensional Poisson equation and once discretised, takes the form of a sparse matrix equation: $\underline{\underline{A}} \underline{p} = \underline{f}$

Despite the sparse nature of $\underline{\underline{A}}$, the inverse $\underline{\underline{A}}^{-1}$ is a dense matrix which reflects the elliptic nature of the problem. Direct evaluation of the inverse $\underline{\underline{A}}^{-1}$, even for a moderately sized grid, is impractical due to the storage of $\underline{\underline{A}}^{-1}$ and to the number of operations required when multiplying $\underline{\underline{A}}^{-1} \underline{f}$. Instead, the sparse matrix equation is solved using a pre-conditioned conjugate gradient algorithm.

The algorithm uses an approximate inverse, $\tilde{\underline{\underline{A}}}^{-1}$ (called the pre-conditioner) to find a correction to the current estimate of the solution. Repeated iteration converges in a finite number of steps to within a finite tolerance, ϵ . The pre-conditioner is used in each iteration. The convergence rate is improved as the pre-conditioner approaches the true inverse, but obviously, the sparser the pre-conditioner is, the fewer the number of computations required per iteration.

A pre-conditioner was designed that is efficient in both storage and number of operations and is a good approximation to $\underline{\underline{A}}^{-1}$. This is accomplished by taking advantage of the geometry of the ocean. The ocean is typically only a few kilometres deep but as much as several thousand kilometres wide. This leads to a dominance in the Laplacian operator, ∇_3^2 , by the vertical contribution, $\frac{\partial^2}{\partial z^2}$. In the sparse matrix corresponding to the 3D Laplacian, just three of the diagonals dominate the seven non-zero diagonals. Retaining just these three diagonals, LU decomposition can be applied rather than storing the inverse of the tri-diagonal. This also reduces the number of computations.

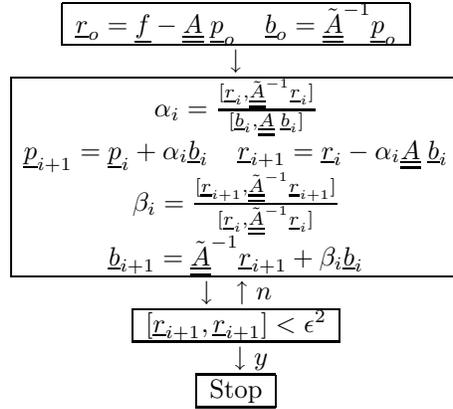


Figure 3.3: Flow diagram for the pre-conditioned conjugate gradient algorithm

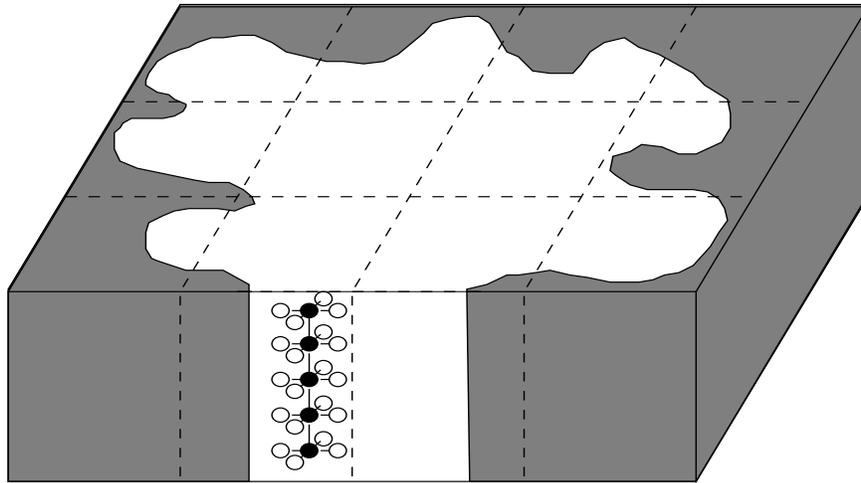


Figure 3.4: Schematic of the ocean domain, decomposed into columns (dashed lines) to which individual processors are dedicated. The shaded regions represent land and the irregular boundaries of the ocean. The stencil represents the local connectivity of the Laplacian operator. The solid dots within the stencil represent the dominant coefficients in the Laplacian and the only connection in the pre-conditioner.

3.8 Computational Aspects of Model: Data parallel architectures

We concentrate our code development on the CM-5, a distributed memory architecture. The code is written in CM-Fortran, a vendor release of Data Parallel Fortran or Fortran 90.

The distributed memory architecture of the CM-5 means that inter-processor communication is an important factor in programming the machine efficiently. The decomposition of arrays into the machines memory can be chosen to reduce the number of cross-processor transfer of information.

The most costly task in the model is the elliptic inversion (whether 2-D or 3-D). The decomposition and the algorithm were both designed to optimize this stage of the model; a significant part of the CPU time is spent in the conjugate gradient algorithm described above. The two principal operators involved here are the Laplacian operator and the pre-conditioner. The Laplacian operator entails nearest neighbour connections between grid points in all three spatial directions (white and black points in figure 3.4). The pre-conditioner connects grid-points only in a vertical line (black points only in figure 3.4). Accordingly, the domain is decomposed into vertical columns, that reach from the top to the bottom of the ocean. This means that vertically aligned points in the model are resident on the same processor thus reducing the communication to zero for the pre-conditioning stage; half of the algorithm.

3.9 Applications of the model

The model has been developed for use at all scales of oceanographic interest. For example, the non-hydrostatic version of the model is applicable to the study of convective overturning. The hydrostatic version of the model can be used to study larger scales right up to the global scale.

By way of illustration of the versatility of the model, two experiments at extreme ends of the spectrum scales are briefly described. The first is a high resolution simulation ($\Delta x = 500\text{m}$) of open ocean convection and the latter is a comparatively low resolution study ($\Delta x \sim 100\text{km}$) of the North Pacific Ocean.

3.9.1 Baroclinic instability of a chimney (convection site)

The non-hydrostatic model is initially at rest with a weak stratification ($N^2 = 5 \times 10^{-8} \text{ s}^{-2}$) in a box $64 \text{ km} \times 64 \text{ km} \times 2 \text{ km}$. The horizontal and vertical resolutions are 500 m and 100 m respectively. Cooling is continuously applied at the surface in a disc of radius 8 km (with a random spatial component). The fluid convectively overturns, fluxing heat upward from the interior to balance the forcing. A mean radial gradient in temperature is established which becomes baroclinically unstable. At some point it becomes more efficient for the fluid to flux heat laterally than vertically so that baroclinic instability begins to dominate.

Figure 3.5 shows the density anomaly at $z=-250\text{m}$ after 10 days. The sharp gradients indicate the front, formed at the edge of the convective zone or chimney, that is undergoing baroclinic instability. This figure is reproduced courtesy of Helen Jones of Legg et al., 1995 [LJVPO].

The model behaves very well in this high resolution limit. There is little grid scale noise. The parallel design of the conjugate gradient algorithm used to solve the 3D elliptic problem is very effective.

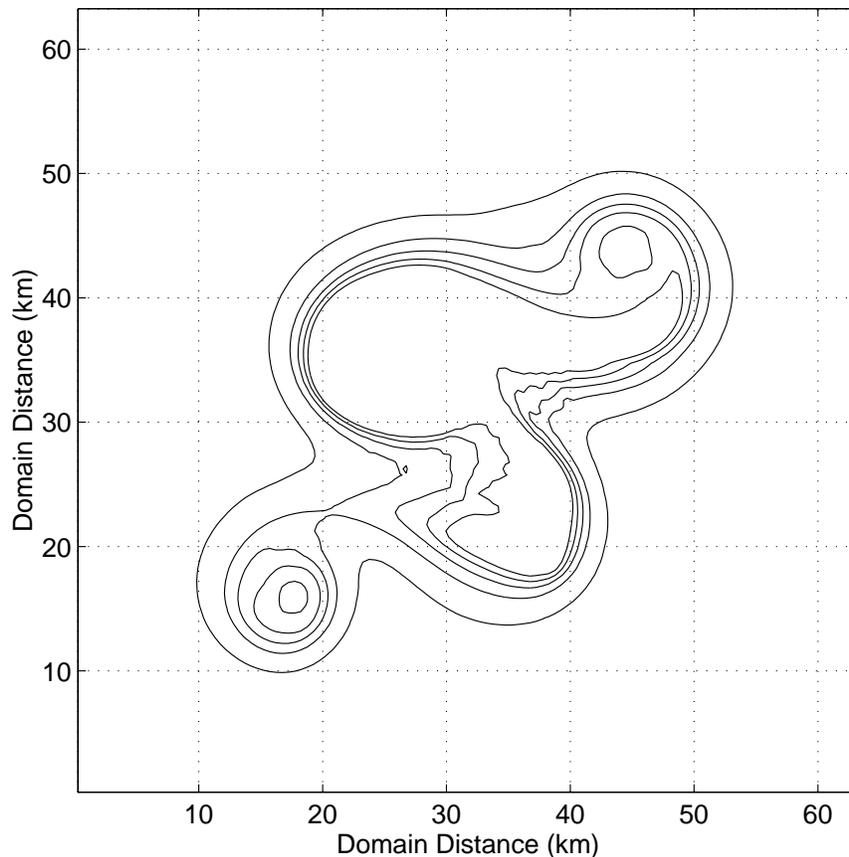


Figure 3.5: Density anomaly at $z=-250\text{m}$ at day 10 in a convection experiment. Reproduced courtesy of Helen Jones (see text).

3.9.2 Climatological spin-up of the North Pacific Ocean

An extended regional calculation is made to obtain a state of the North Pacific Ocean suitable for initializing subsequent integrations. Here, a realistic stratification is desired and so the model is forced with observational data.

The GCM is integrated in hydrostatic mode for the Pacific Ocean (31°S to 61°N, 122°E 293°E). It has realistic coast-lines and bottom topography. Bottom and side walls are insulating. A no-slip (free-slip) condition is used at side walls (bottom). The horizontal grid resolution is 1°. There are 20 levels, with a minimum layer thickness of 25 m at the top to a maximum thickness of 500 m at the bottom.

At the surface, the model is relaxed to climatological values of temperature and salinity [Lev82] with a relaxation time scale of 25 days. At the southern boundary, the relaxation is made over a 5° zone with a time scale of 5 days at the boundary, decreasing linearly to 100 days at 26°S.

The model was initialized from climatological annual mean temperature and salinity distributions [Lev82] and a resting flow field. It was integrated for 17 years with annual mean temperature, salinity and surface wind stress forcing. From year 18 onwards, monthly temperatures, seasonal salinities and monthly winds [TOL89] were used. The forcing fields were updated daily by linearly interpolating from the monthly or seasonal data. Surface heat and freshwater fluxes [Obe88] were introduced in the surface layer starting in year 29. The model time step is 1 hour. The mixing and diffusion coefficients are presented in table 3.1.

Figures 3.6, 3.7 and 3.8 show the state of the model at the end of year 43. The solution exhibits the main features of the observed circulation including the Kuroshio and Oyashio Currents.

This integration used an implementation of the C_d scheme that will be described in chapter 5. Without this innovation, the above integration would not have been possible since the unmodified C grid is susceptible to development of grid scale noise.

To demonstrate the nature of the problem inherent in an unmodified C grid model, the integration was carried on for just one more month after year 43 using the unmodified C grid version of the GCM. Figures 3.9 and 3.10 show the vertical velocity at the end of the extra month. Grid length waves emanate from the boundaries. The waves are predominant at high latitudes where the Rossby radius of deformation is not resolved. With time, the noise propagates through out the model and eventually masks the large scale signal. Note that at the upper level, the ocean is relatively well stratified and that the noise has not reached the interior whilst in the deep, weakly stratified ocean, the noise has penetrated to the Equator.

Horizontal Mixing	ν_h	5×10^3
Vertical Mixing	ν_z	10^{-3}
Horizontal Diffusion	κ_h	5×10^3
Vertical Diffusion	κ_z	3×10^{-5}

Table 3.1: GCM mixing and diffusion coefficients for the North Pacific spin-up. Units are $\text{m}^2 \text{s}^{-1}$.

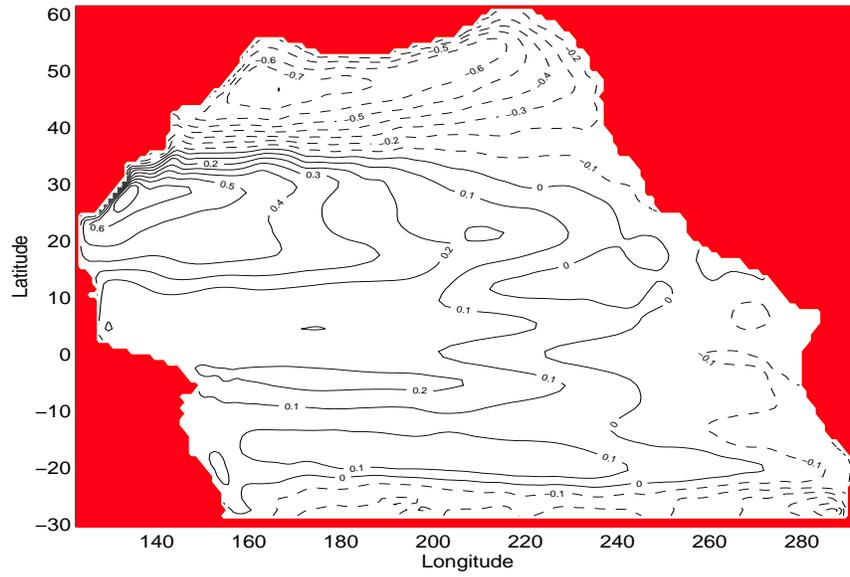


Figure 3.6: Pressure (in m) at $z=-12.5$ m at end of year 43.

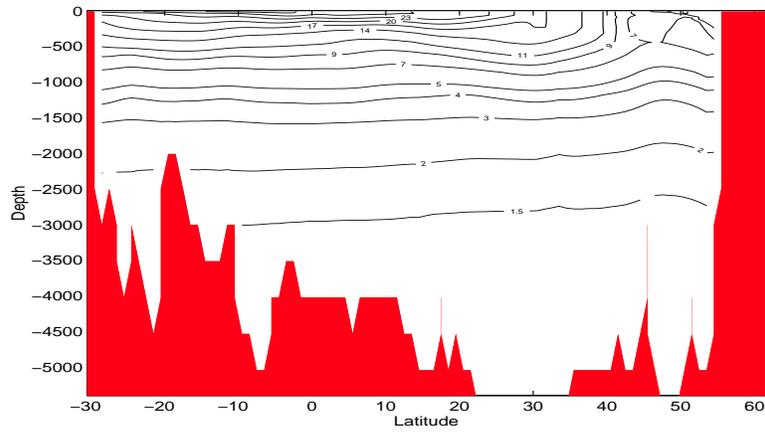


Figure 3.7: Temperature (in $^{\circ}\text{C}$) at $\lambda=171.5\text{E}$ at end of year 43.

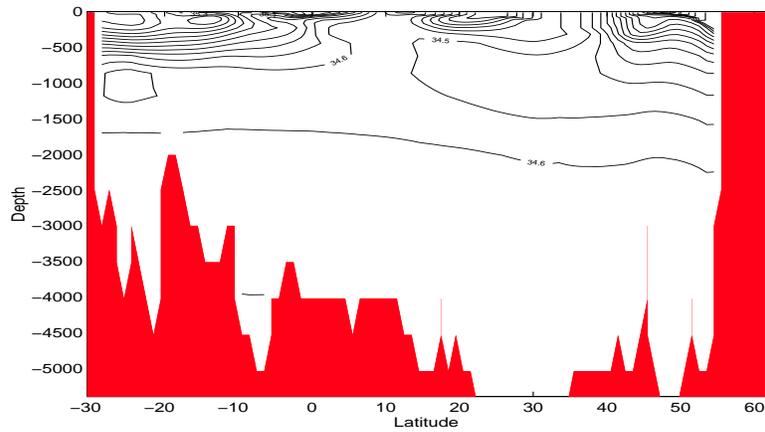


Figure 3.8: Salinity (in psu) at $\lambda=171.5\text{E}$ at end of year 43.

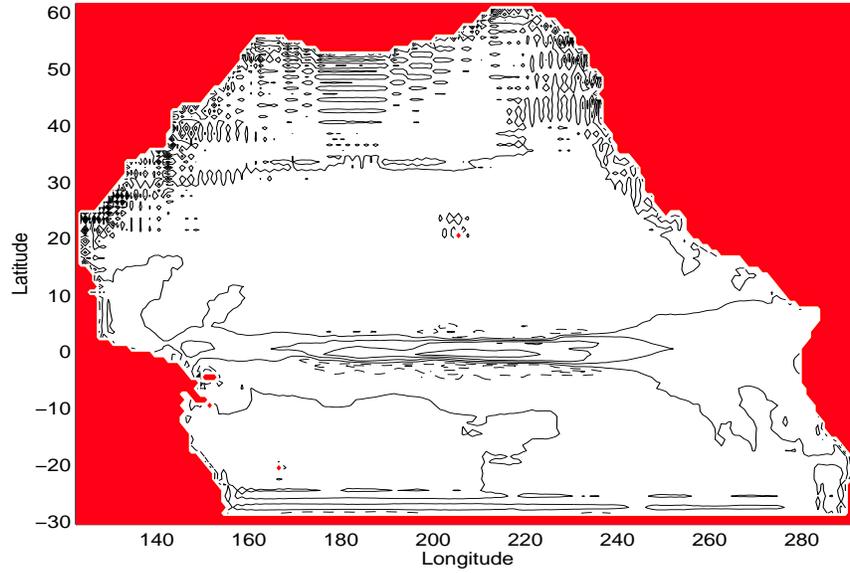


Figure 3.9: Vertical velocity at the base of the top layer in the model after one month of integration with the C_d scheme switched off. Note the presence of grid length waves. Contour interval is 0.01 mm s^{-1} .

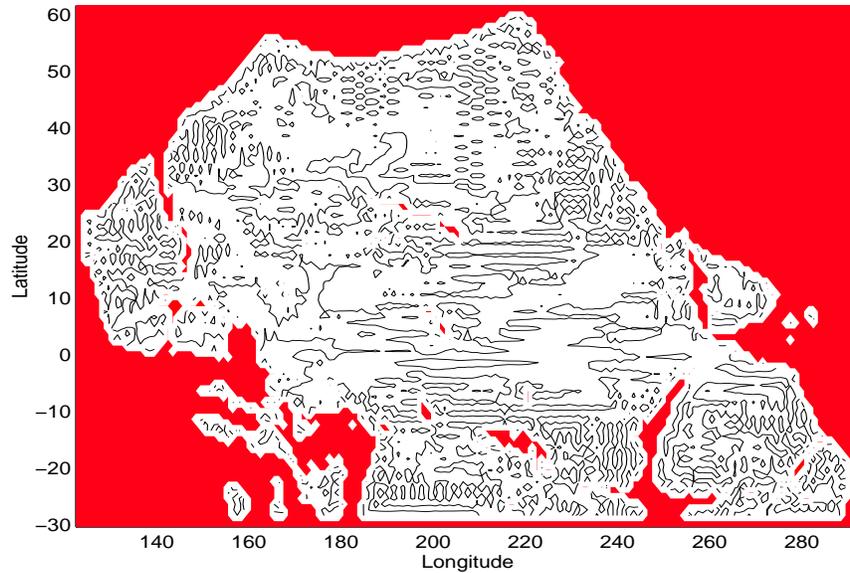


Figure 3.10: Vertical velocity at $z=-3200\text{m}$ in the model after one month. Only the zero contour is shown. The grid length noise is more prevalent at depth because of the weaker stratification.

3.9.3 Conventional treatments of grid scale noise

There have been many attempts in the past to control noise level in ocean models, most frequently in ‘C’ grid formulations that are particularly sensitive to noise at low resolutions.

Several methods that were implemented during the development of the global ocean model are mentioned below but this should by no means be considered a definitive comparison.

Deformation dependent diffusion is a method described by Smagorinsky [Sma63], Deardorff [Dea71] and Smith and Bleck [SBB90] in which the lateral diffusion parameter is a function of the local deformation of the flow. In principle, the deformation field reflects the amplitude of grid-scale noise and acts to smooth the model fields where the noise is manifest. The scheme is physically appealing in that it leaves the physics unaffected except where noise is generated. In practice, a large diffusion coefficient is needed to effectively reduce the noise level and the explicit (in time) evaluation procedure easily violates the numerical stability criteria. An implicit implementation can be employed here but was not investigated.

Biharmonic diffusion This is a commonly used method that is more scale selective than Laplacian diffusion. The method is appealing in that the large scale dynamics is left untouched. In practice again, large coefficients are required but with the limited coefficients allowed for stability (in the explicit model) the method was found to be at least as effective as divergence damping.

Divergence damping This method is based on some understanding of the source of the noise [Xu94]. A non-physical term is added to the horizontal momentum equations; $D_t \mathbf{u} + \dots = \nabla_h \lambda D$ where $D = \nabla_h \cdot \mathbf{u}_h$ is the horizontal divergence. On taking the curl of the momentum equations, it should be apparent that the vorticity dynamics is left unaffected whilst the divergence equation (the poor representation of which is the source of the noise) has an additional diffusion term. This scheme seemed appealing because of this ability to selectively damp noise in the divergence field and in practice seemed to be quite effective in reducing the noise level. A large coefficient, λ , is required but not so large as to cause numerical instability. The only cause for concern was the unfortunate effect at the equator where the dynamics are as much a function of the horizontal divergence field as they are of the vorticity. The sharp upwelling patterns at the equator become very smoothed. In principle, λ could be made a function of latitude to overcome this problem but this was not investigated.

More drastic methods can also be used, such as periodic filtering of the model fields. This would at least be honest in that it deals directly with the noise rather than via the introduction of some “artificial” terms. In practice, none of the above methods work perfectly. They control the problem by reducing the noise level. They all manage this by damping. Wherever there is a source of noise, damping is not good enough. Balancing a source against damping results in a finite amplitude signal.

Instead, the source of the problem should be re-considered. The principle issue here is the gridding. For the ‘C’ grid, the ability to support grid-scale noise is a direct consequence of the spatial averaging required to evaluate the Coriolis terms. Before deriving a method by which this problem is avoided, it is necessary to revisit the theory that describes the grid-noise phenomenon in the model. The next chapter discusses shallow water theory, the propagation of inertia-gravity waves and the finite difference representation of these waves before describing a method that avoids the grid-noise problem.

Chapter 4

Numerical Representation of Inertia-Gravity Waves

Here, we discuss the representation of inertia-gravity waves in finite difference models. The representation of Rossby waves will be considered in the following chapter.

The first sections of the chapter describe Shallow Water theory. It is in this dynamical formulation that the nature of inertia-gravity waves is most accessible. Further more, numerical issues (in particular, the choice of gridding) are traditionally examined in this context.

Having developed an understanding of how the continuous model behaves, an analysis of discrete models on various grids shows how the numerical waves behave as a function of scale. Of particular interest here, is the effect of averaging used in evaluating the Coriolis term on the C grid. The averaging allows grid-scale noise to persist. The noise is manifest in the divergence field and is excited in ocean models by the narrow layers of wind-driven upwelling at boundaries.

The analysis of Arakawa and Lamb, 1977 [AL77], suggests that two particular formulations (the B and C grids using their notation) are most suitable for modeling rotating fluids. To study gridding choices two numerical models were built and integrated in a variety of configurations. The results of the integrations are consistent with the predictions of the numerical analysis.

Whenever models are operating at sub-optimum resolutions, grid-scale noise results and is normally removed by filtering or damping. Instead of dealing with the noise in this crude manner, a philosophy of dealing with the cause of the problem is adopted. Section 4.6 describes a new numerical scheme for treating the Coriolis term on a C grid that accurately represents the inertia-gravity waves and thus avoids the spurious grid-noise problems inherent in the 'C' grid.

The inertia-gravity waves are found to be well behaved at all resolutions using the new scheme. Despite this, there is a new kind of grid-scale noise that appears in the presence of the planetary vorticity gradient, β . The reason for this is that the scheme fails to accurately represent short Rossby waves. This problem will be discussed and dealt with in the next chapter.

4.1 Shallow Water Theory

The dynamics of a homogeneous layer of fluid were briefly analysed in chapter 2. Here two such layers of different density will be considered (see figure 4.1). The two layer system has two intrinsic modes; the barotropic and baroclinic modes. The modes can be separated in the linear system. Although the behavior of linear internal inertia-gravity waves (section 2.4.1) was shown to be quite analogous to the linear external gravity wave discussed previously, it is more appropriate to derive the equations governing the baroclinic mode.

The inviscid, linear, Boussinesq momentum equations on a β -plane for a homogeneous layer are:

$$\frac{\partial u}{\partial t} - f_o(1 + \frac{\beta y}{f_o})v + \frac{1}{\rho_o} \frac{\partial p}{\partial x} = 0 \quad (4.1a)$$

$$\frac{\partial v}{\partial t} + f_o(1 + \frac{\beta y}{f_o})u + \frac{1}{\rho_o} \frac{\partial p}{\partial y} = 0 \quad (4.1b)$$

$$g\rho + \frac{\partial p}{\partial z} = 0 \quad (4.1c)$$

$$\nabla_h \cdot \mathbf{u}_h + \frac{\partial \mathbf{w}}{\partial z} = 0 \quad (4.1d)$$

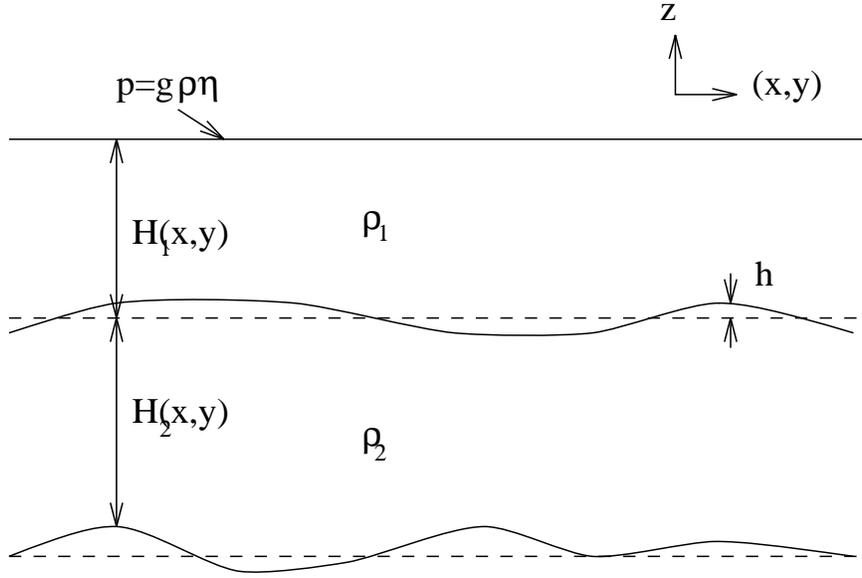


Figure 4.1: The shallow water model comprises of two homogeneous layers of fluid (ρ_1, ρ_2) between rigid-lids, with a free interface of perturbation elevation h . The pressure exerted by the rigid-lid is equivalent to a surface elevation η .

The rigid-lid at the surface exerts a pressure on the fluid of $g\rho_1\eta$, where η is the inferred displacement of the surface. Integrating the hydrostatic relation between some arbitrary depth z and the surface $z = 0$ yields:

$$p(x, y, z) = \begin{cases} p_o + g\rho_1(\eta - z) & \forall h - H_1 < z < \eta \\ p_o + g\rho_1(\eta - h + H_1) + g\rho_2(h - H_1 - z) & \forall H_2 < z < h - H_1 \end{cases} \quad (4.2)$$

and consequently, horizontal gradients of pressure are independent of depth within the layers:

$$\begin{aligned} \nabla p_1 &= \frac{g\rho_1}{\rho_o} \nabla \eta \\ \nabla p_2 &= \frac{g\rho_1}{\rho_o} \nabla(\eta - h) + \frac{g\rho_2}{\rho_o} \nabla h \end{aligned} \quad (4.3)$$

The continuity equation can be applied to each layer to give the rate of change of elevation of the interface:

$$-\frac{\partial h}{\partial t} + \nabla \cdot (H_1 - h)\mathbf{u}_1 = 0 \quad (4.4)$$

$$+\frac{\partial h}{\partial t} + \nabla \cdot (h + H_2)\mathbf{u}_2 = 0 \quad (4.5)$$

These equations can be weighted by $\frac{1}{H_1}$ and $\frac{1}{H_2}$ respectively and added to give:

$$\left(\frac{1}{H_1} + \frac{1}{H_2} \right) \frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}_2 - \mathbf{u}_1) = O\left(\frac{hu_2}{H_2}\right) + \dots \quad (4.6)$$

where the non-linear contributions will be small if the interface perturbation is assumed to be small compared to H_1 and H_2 ($h \ll H_1, H_2$).

Defining:

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{u}_2 - \mathbf{u}_1 \\ g^* &= \frac{g\rho_1}{\rho_o} \\ g' &= \frac{g}{\rho_o}(\rho_2 - \rho_1) \\ H^* &= \frac{H_1 H_2}{H_1 + H_2} \end{aligned}$$

the shallow water equations for the baroclinic mode $\hat{\mathbf{u}}$ can be written:

$$\frac{\partial}{\partial t} \hat{\mathbf{u}} + f_o \left(1 + \frac{\beta y}{f_o}\right) \hat{\mathbf{k}} \wedge \hat{\mathbf{u}} + g' \nabla h = 0 \quad (4.7a)$$

$$\frac{\partial h}{\partial t} + H^* \nabla \cdot \hat{\mathbf{u}} = 0 \quad (4.7b)$$

These equations are prognostic in all three dependent variables which makes the shallow water equations the most simple to integrate numerically (due to the absence of any kind of elliptic inverse problem). For the purpose of analysis, the linearity the equations is also very convenient.

If $H_1 < H_2$ then H^* must fall in the range $\frac{1}{2}H_1 \leq H^* < H_1$. Typically, the bottom layer will be much deeper than the top layer ($H_1 \ll H_2$) and so $H^* \sim H_1$, where a summer time upper-layer depth might be 30 m.

A representative density contrast is $\frac{\rho_2 - \rho_1}{\rho_o} \sim 10^{-3}$ so that $g' \sim 10^{-2}$. The gravity wave speed, which will be derived shortly, is $\sqrt{g'H^*} \sim 0.54 \text{ms}^{-1}$, much slower than the external gravity wave speed $\sqrt{gH} \sim 10 - 100 \text{ms}^{-1}$.

Assuming that \hat{u} and \hat{v} scale in a similar fashion, then the equations can be non-dimensionalised and written in terms of external parameters. The horizontal velocities are non-dimensionalised by a typical velocity scale U . Let the horizontal length scale of the motion be L for a typical interface displacement of h_o and the time-scale be f_o^{-1} then the non-dimensional equations may be written:

$$\frac{\partial}{\partial t} \mathbf{u} + (1 + \beta y) \hat{\mathbf{k}} \wedge \mathbf{u} + \frac{g' h_o}{f_o L U} \nabla h = 0 \quad (4.8a)$$

$$\frac{f_o^{-1} h_o L}{H^* U} \frac{\partial h}{\partial t} + \nabla \cdot \mathbf{u} = 0 \quad (4.8b)$$

where all the variables are now non-dimensional. Here, β is the non-dimensional planetary vorticity gradient, $\beta \rightarrow \frac{\beta L}{f_o}$.

The non-dimensional quantity in the momentum equations can be expressed:

$$\frac{g' h_o}{f_o L U} = \frac{g' H^*}{U^2} \frac{h}{H^*} \frac{U}{f_o L} = R_i \delta R_o \quad (4.9)$$

where $R_i = \frac{g' H^*}{U^2} = \frac{c_g^2}{U^2}$ is the Richardson number and $R_o = \frac{U}{f_o L}$ is the Rossby number. $\delta = \frac{h}{H^*}$ is a small parameter measuring the magnitude of perturbation elevation against the vertical scale height H^* .

Similarly, the non-dimensional number appearing in the continuity equation can be expressed:

$$\frac{f_o^{-1} h_o L}{H^* U} = \frac{h_o}{H} \frac{f_o L}{U} = \delta R_o^{-1} \quad (4.10)$$

Since, the time-scales of interest are longer than f_o^{-1} , then for the non-dimensional velocity to be order one, the pressure gradient term should be order one also. The internal parameter δ can therefore be chosen:

$$\delta R_i R_o = 1 \quad (4.11)$$

The non-dimensional shallow water equations are then:

$$\frac{\partial}{\partial t} \mathbf{u} + (1 + \beta y) \hat{\mathbf{k}} \wedge \mathbf{u} + \nabla h = 0 \quad (4.12a)$$

$$S^{-1} \frac{\partial h}{\partial t} + \nabla \cdot \mathbf{u} = 0 \quad (4.12b)$$

where the non-dimensional parameter $S = R_i R_o^2 = \frac{g' H^*}{f_o^2 L^2} = \frac{L^2}{L_g^2}$ is the Burger number, which is the square of the ratio between the Rossby deformation radius and the length scale. This ratio will prove to be very relevant when an analysis of the numerical models is made. The crucial quantity will be the Burger number of the grid-scale, better termed the wave resolution.

The kinetic and potential energy take the forms $KE = \frac{1}{2} \mathbf{u} \cdot \mathbf{u}$, $PE = \frac{1}{2} S^{-1} h^2$. Total energy, $\iint KE + PE dA$, is conserved in the absence of external forces.

The shallow water equations derived here are analogous to the barotropic equations describing the external gravity mode. The system exhibits all of the properties pertinent to an individual baroclinic mode of the hydrostatic primitive equations. This model was derived for this property so that it can be used as a elementary setting to analyse the numerical models.

4.2 Inertia-Gravity Waves

The Coriolis terms in isolation act to make the flow move in inertial circles. The pressure gradient terms allow a pressure anomaly to propagate as a gravity wave. In conjunction, both motions are described by a pair of inertia-gravity wave modes as follows. The divergence, vorticity and continuity equations are:

$$\frac{\partial}{\partial t}D - \zeta + \nabla^2 h = 0 \quad (4.13a)$$

$$\frac{\partial}{\partial t}\zeta + D = 0 \quad (4.13b)$$

$$S^{-1}\frac{\partial}{\partial t}h + D = 0 \quad (4.13c)$$

The inertia-gravity waves are best described in terms of the divergence field. Differentiating the divergence equation and eliminating the vorticity and pressure terms yields a wave equation of the form:

$$\frac{\partial^2}{\partial t^2}D + D - S\nabla^2 D = 0 \quad (4.14)$$

Substituting a plain wave solution of the form $e^{i(kx+ly-\omega t)}$ into the wave equation yields the dispersion relation for the inertia-gravity waves:

$$\omega^2 = 1 + S(k^2 + l^2) \quad (4.15)$$

The frequency, ω , of a gravity wave is thus a monotonically increasing function in $|\boldsymbol{\kappa}| = \sqrt{k^2 + l^2}$, with a minimum frequency of $\omega = 1$ corresponding to $S = 0$ or $\boldsymbol{\kappa} = 0$. For long waves, $S|\boldsymbol{\kappa}|^2 \ll 1$, the frequency is approximately $\omega \approx 1$ and so the dominant frequency is set by the Coriolis acceleration. For short waves, $S|\boldsymbol{\kappa}|^2 \gg 1$, the frequency becomes proportional to the wave number, $\omega \approx \sqrt{S}|\boldsymbol{\kappa}|$. In this limit, the pressure gradient terms are the predominant ones controlling the wave propagation.

The group velocity, $\frac{\partial \omega}{\partial \mathbf{k}}$ is:

$$\mathbf{c}_g = \frac{S}{\sqrt{1 + S|\boldsymbol{\kappa}|^2}} \begin{pmatrix} k \\ l \end{pmatrix}$$

and always points in the direction of the wave vector, $\boldsymbol{\kappa}$. For the long waves the group speed is proportional to the wave number $c_g \approx S\boldsymbol{\kappa}$ while for short waves the group speed asymptotes to $c_g = \sqrt{S}\frac{1}{|\boldsymbol{\kappa}|}(k, l)$. Never does the group speed approach zero for non-zero k or l .

This, then, is how linear inertia gravity waves propagate in the continuum. Although a discrete representation cannot be expected to reproduce this behavior perfectly, it is useful to examine the dispersion relation and group speed on the grid to assess the closeness of the numerical representation of the dynamics to that of the real ocean and so qualitatively gauge the numerical scheme. A celebrated account of this analysis was given by Arakawa and Lamb [AL77] which will be summarised later.

4.3 Damped wave motion

The inertia-gravity wave motions just described allow the fluid to adjust to a geostrophic state. In an inviscid model, the gravity waves travel and reflect off boundaries, unimpaired by diffusion. Such a model would never reach a steady state. Dissipation of some form must be represented. Accordingly, an arbitrary operator, \hat{F} , will be introduced to represent the dissipation. Damping of the continuity equation can also be introduced, \hat{H} , which reflects the dissipation in the buoyancy equation of the primitive equations.

$$\frac{\partial}{\partial t}\mathbf{u} + \hat{\mathbf{k}} \wedge \mathbf{u} + \nabla h = \hat{F}(\mathbf{u}) \quad (4.16a)$$

$$\frac{\partial}{\partial t}h + S\nabla \cdot \mathbf{u} = \hat{H}(h) \quad (4.16b)$$

As before, the three dependent variables can easily be exchanged for two dependent variables by expressing the equations in terms of divergence, vorticity and elevation, and then eliminating the vorticity:

$$\left(\frac{\partial}{\partial t} - \hat{F}\right)^2 D + D + \left(\frac{\partial}{\partial t} - \hat{F}\right)\nabla^2 h = 0 \quad (4.17a)$$

$$\left(\frac{\partial}{\partial t} - \hat{H}\right)h + SD = 0 \quad (4.17b)$$

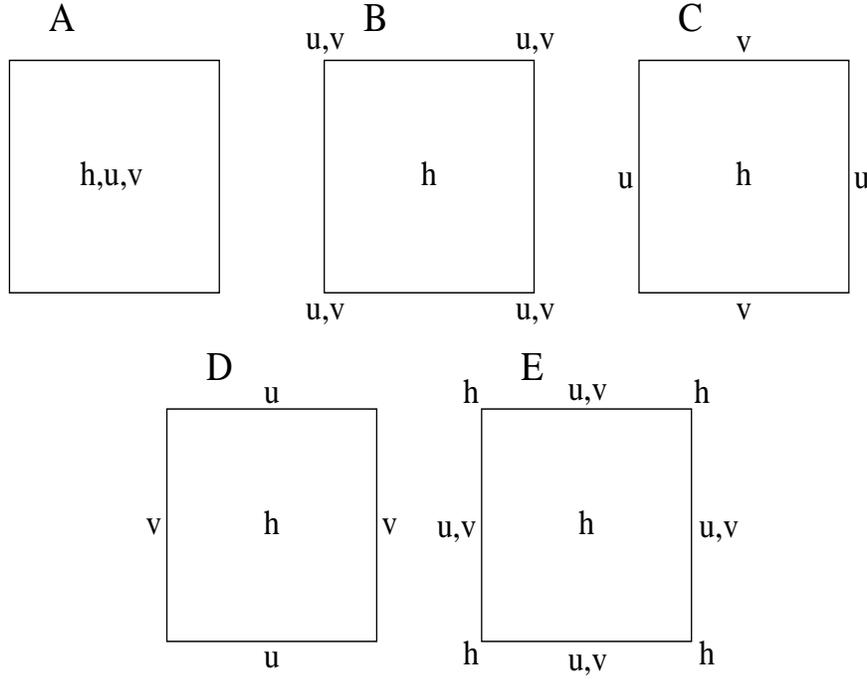


Figure 4.2: The variables of a model can be arranged in only five grids that maintain second order accuracy. Grids A-D share the same grid-spacing. Grid E is equivalent to the B grid rotated through 45° and scaled by $\sqrt{2}$.

In the special case where $\hat{H} = \hat{F}$, the gravity wave motion is easily found by eliminating h :

$$\left(\frac{\partial}{\partial t} - \hat{F}\right)^2 D + D + S\nabla^2 D = 0 \quad (4.18)$$

which gives the dispersion relation:

$$\omega = i\hat{F} \pm (1 + S(k^2 + l^2)) \quad (4.19)$$

The real, oscillatory part of ω is just the plain inertia-gravity wave contribution from the inviscid model; $\pm(1 + S(k^2 + l^2))$. The imaginary, decaying part is simply the damping rate of the operator for that particular wave number. The inertia-gravity waves are damped, but the phase and group speeds of the oscillatory solution remain unchanged.

4.4 Finite-differenced Inertia-gravity waves

Using centered finite differencing and the simplest single-level time stepping scheme, the shallow water equations can be discretised on five distinct grids (see figure 4.2). Arakawa and Lamb [AL77], for the purpose of analysis, assumed that the time-stepping scheme was perfect and retained the continuous time derivative. It is in fact quite simple to consider discrete time-stepping but as long as the scheme is centered (and thus second order accurate), no new consequences arise.

Scheme A:

$$\delta_t u - \bar{v}^t + \delta_x \bar{h}^x = 0 \quad (4.20a)$$

$$\delta_t v + \bar{u}^t + \delta_y \bar{h}^y = 0 \quad (4.20b)$$

$$S^{-1} \delta_t h + (\delta_x \bar{u}^x + \delta_y \bar{v}^y) = 0 \quad (4.20c)$$

Scheme B:

$$\delta_t u - \bar{v}^t + \delta_x \bar{h}^y = 0 \quad (4.21a)$$

$$\delta_t v + \bar{u}^t + \delta_y \bar{h}^x = 0 \quad (4.21b)$$

$$S^{-1} \delta_t h + (\delta_x \bar{u}^y + \delta_y \bar{v}^x) = 0 \quad (4.21c)$$

Scheme C:

$$\delta_t u - \overline{v}^{xyt} + \delta_x h = 0 \quad (4.22a)$$

$$\delta_t v + \overline{u}^{xyt} + \delta_y h = 0 \quad (4.22b)$$

$$S^{-1} \delta_t h + (\delta_x u + \delta_y v) = 0 \quad (4.22c)$$

Scheme D:

$$\delta_t u - \overline{v}^{xyt} + \delta_x \overline{h}^{xy} = 0 \quad (4.23a)$$

$$\delta_t v + \overline{u}^{xyt} + \delta_y \overline{h}^{xy} = 0 \quad (4.23b)$$

$$S^{-1} \delta_t h + (\delta_x \overline{u}^{xy} + \delta_y \overline{v}^{xy}) = 0 \quad (4.23c)$$

The E grid is omitted since it is equivalent to the B grid rotated through 45° and scaled by $\sqrt{2}$.

The divergence equations for each model are as follows:

A grid:

$$D \equiv \delta_x \overline{u}^x + \delta_y \overline{v}^y \quad \delta_{tt} D + \overline{D}^{tt} - S(\delta_{xx} \overline{D}^{xx} + \delta_{yy} \overline{D}^{yy}) = 0 \quad (4.24)$$

B grid:

$$D \equiv \delta_x \overline{u}^y + \delta_y \overline{v}^x \quad \delta_{tt} D + \overline{D}^{tt} - S(\delta_{xx} \overline{D}^{yy} + \delta_{yy} \overline{D}^{xx}) = 0 \quad (4.25)$$

C grid:

$$D \equiv \delta_x u + \delta_y v \quad \delta_{tt} D + \overline{D}^{xxyytt} - S(\delta_{xx} D + \delta_{yy} D) = 0 \quad (4.26)$$

D grid:

$$D \equiv \delta_x u + \delta_y v \quad \delta_{tt} D + \overline{\overline{D}^{xxyy}}^{tt} - S(\delta_{xx} D + \delta_{yy} D) = 0 \quad (4.27)$$

Substitution of a plane wave solution ($\exp i(kx + ly - \omega t)$) into the divergence equations yields:

$$\text{A: } \omega^2 = 1 + S(s_k^2 c_k^2 + s_l^2 c_l^2) \quad (4.28a)$$

$$\text{B: } \omega^2 = 1 + S(s_k^2 c_l^2 + s_l^2 c_k^2) \quad (4.28b)$$

$$\text{C: } \omega^2 = c_k^2 c_l^2 + S(s_k^2 + s_l^2) \quad (4.28c)$$

$$\text{D: } \omega^2 = c_k^2 c_l^2 + S c_k^2 c_l^2 (s_k^2 + s_l^2) \quad (4.28d)$$

where the abbreviations, $s_k = \frac{2}{\Delta x} \sin k \frac{\Delta x}{2}$, $s_l = \frac{2}{\Delta y} \sin l \frac{\Delta y}{2}$, $c_k = \cos k \frac{\Delta x}{2}$ and $c_l = \cos l \frac{\Delta y}{2}$ have been used for brevity. In the limit of infinitely high spatial resolution, $c_k \rightarrow 1$ and $s_k \rightarrow k$, etc. Thus all the finite difference dispersion relations approach the continuous limit though they may still contain extra turning points.

The common term $S s_k^2 = \frac{4S}{\Delta x^2} \sin^2 k \Delta x / 2$ is the discrete analogue of $S k^2$. The quantity $4S/\Delta x^2$ is the wave resolution parameter, which appears in all the dispersion relations. It is the square of the ratio between twice the dimensional Rossby deformation radius and the dimensional grid-spacing. If $4S/\Delta x^2 = 1$, the grid spacing is small enough to resolve a single wavelength equal to the deformation radius. High resolution corresponds to $4S/\Delta x^2 \gg 1$ and low resolution to $4S/\Delta x^2 \ll 1$.

Figure 4.4 shows the non-dimensional frequency as a function of wave resolution, $m = k\Delta x/\pi$, $n = l\Delta y/\pi$ for high resolution models on the four different grids. For comparison, the analytic dispersion relation for the continuous system is plotted in figure 4.3 for the same mode numbers.

The gradient of the surfaces is the group velocity and should always be directed away from the origin. At high resolution, the discrete dispersion relations might be expected to approach the analytic result as the resolution increases. This is always found but the discrete cases always contain extra minima at the highest wave numbers.

The A grid has three erroneous minima occurring at the two grid-scale wave length; $m = 1$, $n = 1$ and $m = n = 1$. Between these points and the turning points ($m = 1/2$ and $n = 1/2$) the group velocity has the wrong sign. The phase speed is underestimated because the frequency of motion is much lower than it should be.

The B grid has one extra minima for the checker board mode (grid-scale mode number in each direction, $m = n = 1$). The group velocity has the wrong sign for short waves that satisfy $m + n > 1$. Again, the phase speed and frequency are severely underestimated but this time only in one quadrant.

The C grid underestimates the frequency of the highest modes. In the analytic case, the frequency reached approximately 2.5 cycles whilst the peak C grid frequency is about 2 cycles. The group speed is always of the right sign with the highest errors at high wave numbers.

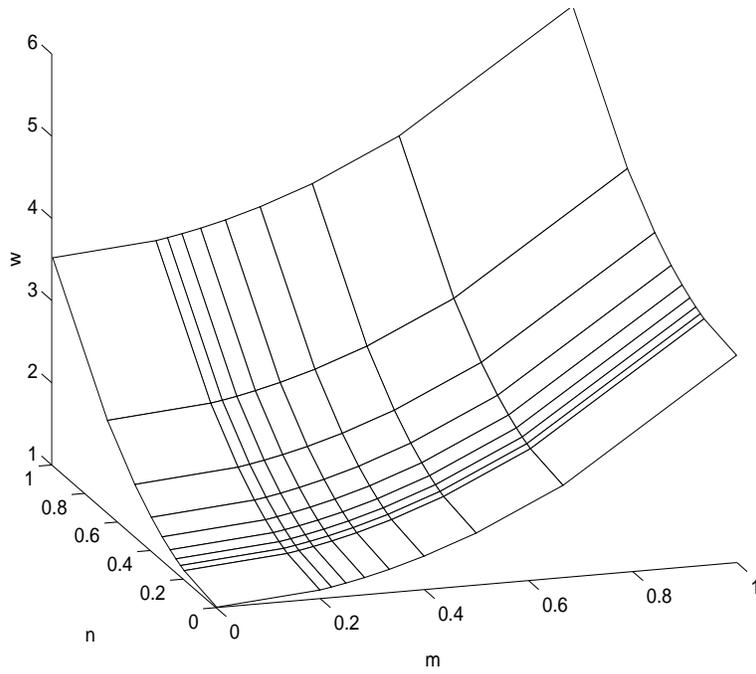


Figure 4.3: Non-dimensional frequency of the continuous spectrum is plotted against non-dimensional wave number $\{m, n\}$. The discrete wave numbers allowed in the model are represented by the grid-lines. The wave numbers correspond to the next figure where $\frac{4S}{\Delta x^2} = 2$.

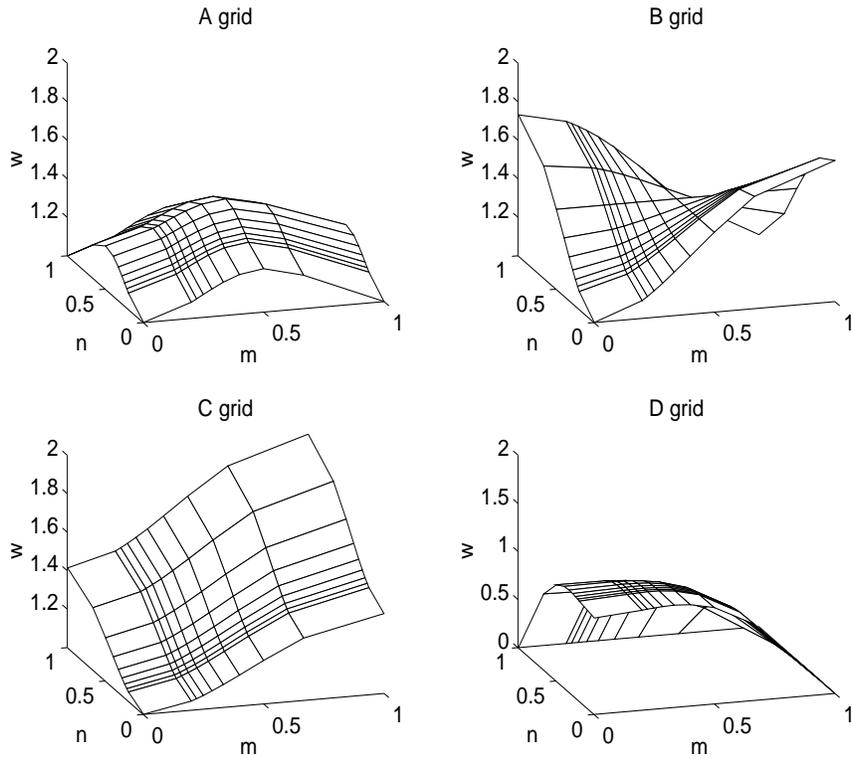


Figure 4.4: Non-dimensional frequency of the discrete waves is plotted against wave-resolution $\{m, n\}$. The Rossby radius is resolved, $\frac{4S}{\Delta x^2} = 2$. The discrete wave numbers allowed in the model are represented by the grid-lines.

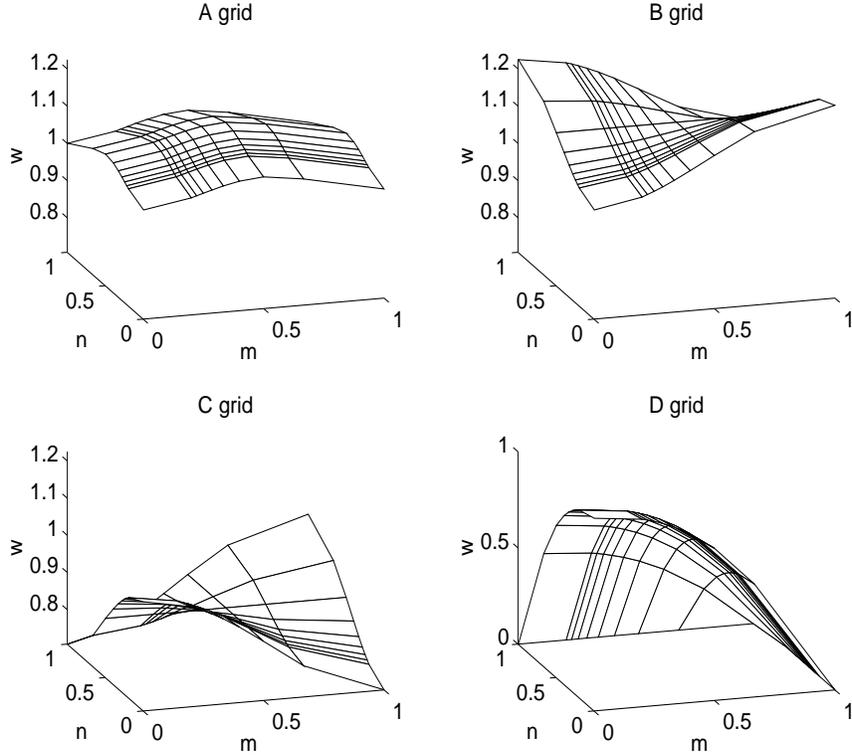


Figure 4.5: Non-dimensional frequency of the discrete waves is plotted against wave-resolution $\{m, n\}$. The Rossby radius is not resolved, $\frac{4S}{\Delta x^2} = \frac{1}{2}$. The discrete wave numbers allowed in the model are represented by the grid-lines.

The D grid exhibits the worst performance, having a minimum value of zero frequency for any grid scale mode ($m = 1$ or $n = 1$). The group velocity behaves much like in the A grid, except that for the higher modes it is considerably more negative, and thus more erroneous.

The conclusion one should draw from these four dispersion relations is that at high resolution, the C grid is the only grid that represents energy propagation for all scales in the correct qualitative manner. On all other grids, the group velocity points in the wrong direction and energy may accumulate.

Figure 4.5 shows the dispersion relation for the four models at low resolution ($4S/\Delta x^2 = \frac{1}{2}$). The A, B and D grid models behave in the same qualitative manner as before, though now, because the range of frequencies is much reduced, the fractional error in group speed and frequencies is similarly reduced.

The C grid model has changed its behavior. There are two extra minima corresponding to waves that are long in one direction, short in the other. The group speed is pointed in the wrong direction for long waves. This last factor renders the model much less accurate than other models since it is the large scale features which are erroneously affected.

In summary, at high resolution, where the geostrophic adjustment process is resolved this analysis clearly suggests that the ‘C’ grid is the most suitable grid. At low resolution, the C grid has severe problems and the ‘B’ grid seems to be the best compromise. The error in the B grid frequency for high mode numbers is less significant at low resolution because of the reduced range in frequencies.

The failing of the ‘C’ grid at low resolution is a consequence of spatial averaging of the Coriolis term which introduces the cosines into the dispersion relations (4.28) for the gravity waves. The effect of the averaging operator is to filter out the two grid-length wavelengths which leads to a vanishing of the Coriolis term in (4.28).

The invisibility of a two grid-length wave to the averaging operator is easily illustrated. Consider the spatial pattern ... +1 -1 +1 -1 +1 ... to represent the two grid-length wave on a regularly spaced grid. Apply a spatial average to obtain the value at the mid-points. The result is simply zero. For the ‘C’ grid, a two grid-length wave of arbitrary amplitude can be added to a component of the flow without affecting the discrete evaluation of the Coriolis term. Moreover, this wave remains fixed in space (zero group speed) and so may grow. Similarly, in the case of the ‘B’ grid, a checker board type pattern in the pressure field is ‘invisible’ to the pressure gradient term.

In summary then, the preferred grid depends on its resolution compared to the Rossby deformation radius. If

this length scale is not resolved then the shortest resolved scales of motion are aware of the Coriolis effect and so the 'B' grid is the better choice. If this length scale is resolved, then the shortest motions will be gravity waves and the dominant process controlling their evolution will be the pressure gradient term and so the 'C' grid will be the better choice.

This criterion works well if the deformation radius is relatively fixed, both in time and space. However, in baroclinic models of the atmosphere and oceans, the deformation radius is a function of latitude, the local stratification and the vertical wave number of the particular baroclinic mode, $L_D^{(m)} = N/f(\phi)m^{-1}$, where m is the dimensional vertical wave number, $f(\phi)$ is the local Coriolis parameter and N is the local Brunt-Vasala frequency. Thus, with a fixed grid, there is the possibility of both resolving and not resolving the local deformation radius in separate parts of a model, or for different vertical modes.

In the case of the GCM described in chapter 3, the model is used to study the global scale circulation. This demands a coarse resolution because of the limitations of the computer. Hence a 'B' grid might be argued for. Here, however, the model is also used to study convection which demands the the formulation on a C grid.

4.5 Numerical Shallow Water models

Having developed a linear tool to study grid-scale noise in the GCM, it remains to test the ideas in practice. Since only two models, the 'B' and 'C' grids, are expected to be reliably accurate (depending on wave resolution), these were the ones developed.

The shallow water models described earlier were second order accurate in time. This meant that the evaluation of the Coriolis term was made in a semi-implicit manner. Implicit Coriolis terms on the A and B grids involves simple algebraic manipulation of the discretised equations. On the C and D grids, because of the spatial averaging, implementing implicit Coriolis terms involves finding the inverse of a matrix describing the spatial dependencies.

In order to compare models to examine the effects of gridding choice, these models should use the same algorithms wherever possible in order to isolate the cause of any differences. The C grid restricts the modeller to an explicit evaluation of the Coriolis terms (for a reasonably efficient code). In all of the following models, the quasi-second order Adams-Bashforth scheme is applied to evaluate the Coriolis terms.

The B and C grid models will now be described. The same grid resolutions are used in both whenever the models are compared. The grid resolution is defined by the number of pressure points in each direction. Thus, the pressure field is carried at the same physical locations in each model. Both models carry the normal velocity on the physical boundaries. The B grid also carries the tangential velocity on the boundary and an extra boundary condition is thus needed here. The simplest choice is that of no-slip.

The models are forced with an analytically described wind stress. Dissipation is parameterised by a linear friction like term with time-scale λ^{-1} . Had a higher order diffusion term been chosen, the grid-scale noise would be more efficiently removed and the discrepancies between the two grids less obvious.

4.5.1 B grid shallow water model

The grid of the B grid model is defined by cells centered about the pressure point, h . For $M \times N$ pressure points, there are therefore $(M + 1) \times (N + 1)$ points that carry the velocities u and v , although $2(M + N)$ of these lie on the boundary.

The time-stepping cycle is comprised of two prognostic steps; one stepping forward the continuity equation and the other stepping forward the momentum equations.

The continuity equation is time differenced as described before; single time level with a centered in time and space evaluation of the divergence field:

$$h^{n+\frac{1}{2}} = h^{n-\frac{1}{2}} + \Delta t S (\delta_x \overline{u^{n,y}} + \delta_y \overline{v^{n,x}}) \quad (4.29)$$

No flux through solid boundaries is a necessary condition to evaluate the right hand side. The initial state is one of no motion.

The next stage in the model time-step is to step forward the momentum equations as follows:

$$u^{n+1} = u^n + \Delta t \left(+(1 + \epsilon_{\beta y})v^{(n+\frac{1}{2})} - \delta_x \overline{h^{n+\frac{1}{2}}{}^y} + \tau^x - \lambda u^n \right) \quad (4.30a)$$

$$v^{n+1} = v^n + \Delta t \left(-(1 + \epsilon_{\beta y})u^{(n+\frac{1}{2})} - \delta_y \overline{h^{n+\frac{1}{2}}{}^x} + \tau^y - \lambda v^n \right) \quad (4.30b)$$

where the parenthesized superscript $(n + \frac{1}{2})$ denotes the Adams-Bashforth extrapolation: $u^{(n+\frac{1}{2})} = (\frac{3}{2} + \epsilon_{AB})u^n - (\frac{1}{2} + \epsilon_{AB})u^{n-1}$. No flow normal to solid boundaries is imposed in the continuity equation.

	Low resolution	High resolution
Burger number S	1.0×10^{-4}	6.25×10^{-2}
Planetary vorticity gradient β	0.0	0.0
Friction λ	3×10^{-2}	3×10^{-2}
Time step Δt	0.2	0.2
Adams-Bashforth parameter ϵ_{AB}	0.1	0.1
Grid spacing $\Delta x = \Delta y$	1/10	1/10
Wave resolution $2\sqrt{S}/\Delta x$	1/5	5

Table 4.1: External parameters used in the comparison of the B and C grid shallow water models. High resolution means that the Rossby deformation radius is resolved whilst Low resolution means that it is not.

4.5.2 C grid shallow water model

As in the B grid model, the C grid is defined by cells centered about the pressure point, h . For $M \times N$ pressure points, there are therefore $(M + 1) \times N$ points that carry the zonal velocity, u , and $M \times (N + 1)$ points that carry the meridional velocity, v .

The time-stepping cycle is as for the B grid. The continuity equation is time differenced with a single time level with a centered in time and space evaluation of the divergence field:

$$h^{n+\frac{1}{2}} = h^{n-\frac{1}{2}} + \Delta t S (\delta_x u^n + \delta_y v^n) \quad (4.31)$$

The vanishing of the normal flux through solid boundaries is applied quite naturally.

The momentum equations are stepped forward as follows:

$$u^{n+1} = u^n + \Delta t \left(+(1 + \epsilon_{\beta y}) \overline{v^{(n+\frac{1}{2})} }^{xy} - \delta_x h^{n+\frac{1}{2}} + \tau^x - \lambda u^n \right) \quad (4.32a)$$

$$v^{n+1} = v^n + \Delta t \left(-(1 + \epsilon_{\beta y}) \overline{u^{(n+\frac{1}{2})} }^{xy} - \delta_y h^{n+\frac{1}{2}} + \tau^y - \lambda v^n \right) \quad (4.32b)$$

where the parenthesized superscript (n) again denotes the Adams-Bashforth extrapolation: $u^{(n)} = (\frac{3}{2} + \epsilon_{AB})u^n - (\frac{1}{2} + \epsilon_{AB})u^{n-1}$.

4.5.3 Shallow water model results

The B and C grid models described above were configured to mimic a double gyre on an f-plane. The model extents were $x : \{0 \leq x \leq 1\}$ and $y : \{-1 \leq y \leq 1\}$. The applied wind stress was circular and non-divergent, $\tau = \hat{k} \wedge \nabla(\sin \pi x \sin \pi y)$ where the appropriate discrete operator was used to evaluate the curl.

The external parameters of the models are listed in table 4.1. The Burger number is defined using the physical zonal extent of the basin, $S = L_D^2/L^2$. The wave resolution is thus $2\sqrt{S}/\Delta x$ and was 5 for the high resolution run and 1/5 for the low. The symmetric forcing and absence of β effect ($\epsilon_{\beta} = 0$) means that the solutions should be symmetric.

Low resolution: Figures 4.6 and 4.7 show the low resolution B and C grid models at $t = 100$. The wave resolution is $2\sqrt{S}/\Delta x = 1/5$. At this resolution the B grid reproduces the analytical solution very well. The pressure field and divergence patterns in the B grid both reflect the structure of the forcing. The C grid however is clearly being influenced by boundary effects. The large values of upwelling and downwelling at the boundaries alternate as one moves from the boundary. In the corners, the alternating zonal and meridional patterns interfere to produce a checker-board pattern. All three of the short wave combinations (meridional, zonal, checker-board) correspond to the minimum frequencies in the dispersion relation of the C grid. The erroneous pressure field in the C grid is also driving fast flows which bear no relation to the more reasonable circulation of the B grid.

These integrations mimic the previous low resolution integrations of the GCM. The grid-scale noise is similar in character. The conclusion here is that a GCM developed on a B grid would perform better at low resolution.

High resolution: The high resolution integrations tell a different tale. Figures 4.8 and 4.9 show the high resolution B and C grid models at $t = 100$. The resolution is now $2\sqrt{S}/\Delta x = 5$. The pressure field and circulation both appear

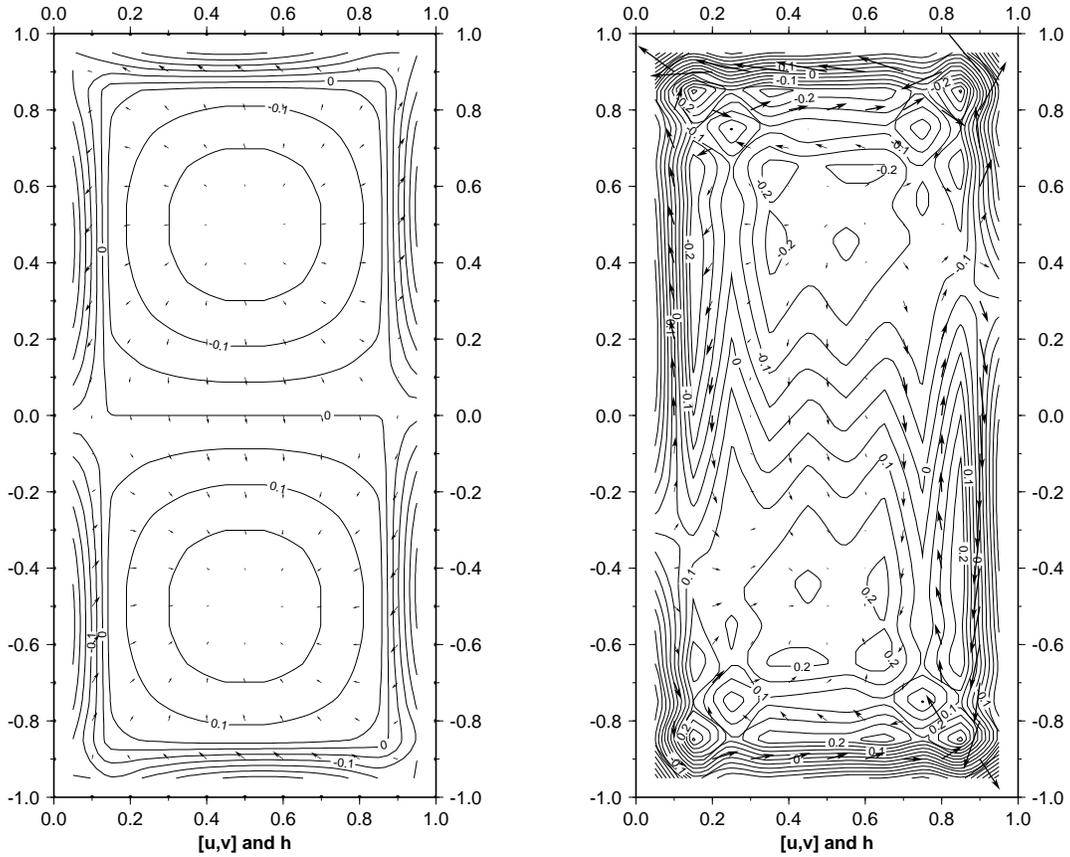


Figure 4.6: Low resolution ($4S/\Delta x^2 = 1/5$) integrations of the B grid (left) and C grid (right) shallow water models. Shown are pressure (contours) and velocities (arrows) at time $t = 100$. Contour interval for pressure is 0.05. Note that the models are still spinning up and so the flow is ageostrophic. At this resolution, the IG waves should behave better in the B grid than in the C grid. The B grid pressure field reflects the structure of the forcing whilst the C grid is being disturbed by boundary effects.

to be symmetric and consistent with each other. The B grid pressure maximum is a few percent weaker than the C grid, the latter being the more reliable model at this resolution.

The divergence patterns are, however, very different. The B grid has a very strong checker-board pattern that is an order of magnitude larger than the large scale divergence pattern of the C grid. Only the checker-board mode seems to have been excited, as predicted by the minima in the B grid dispersion relation.

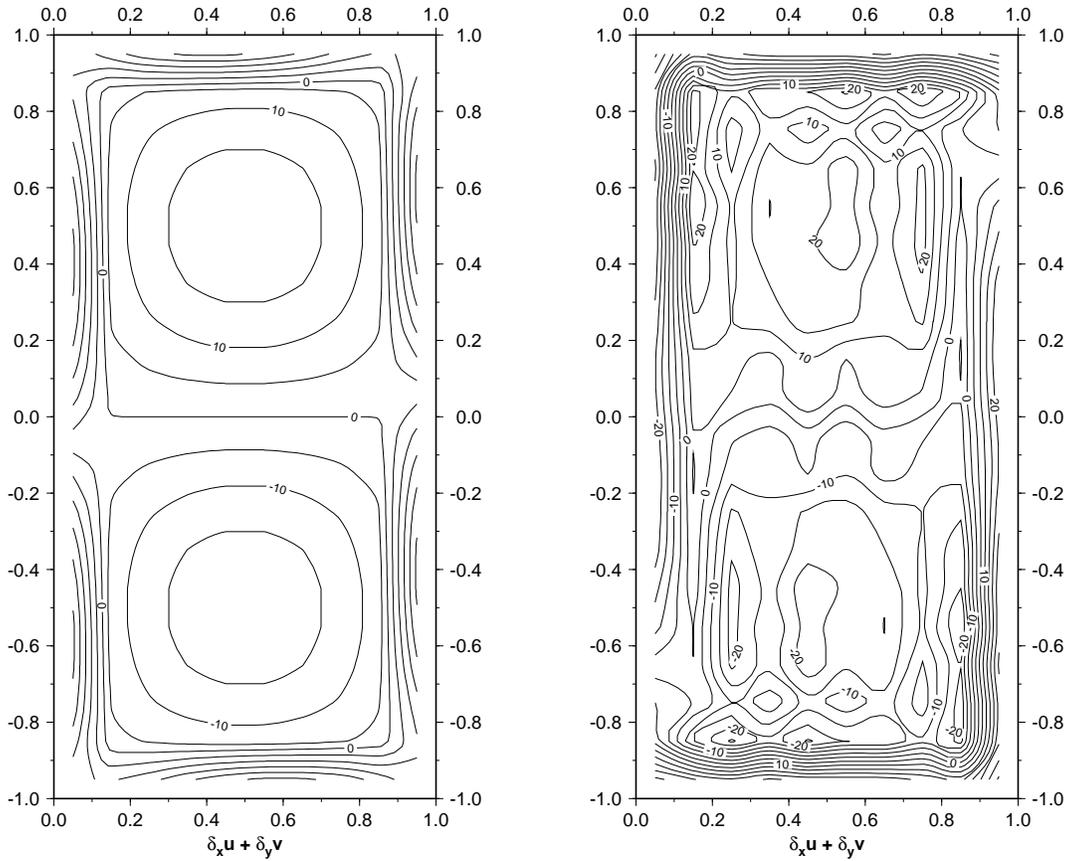


Figure 4.7: Low resolution ($4S/\Delta x^2 = 1/5$) integrations of the B grid (left) and C grid (right) shallow water models. Shown is the horizontal divergence at time $t = 100$. Contour interval is 5. The flow is ageostrophic at this stage in the integration so the divergence is still large. The C grid clearly has more signal at the grid scale than does the B grid.

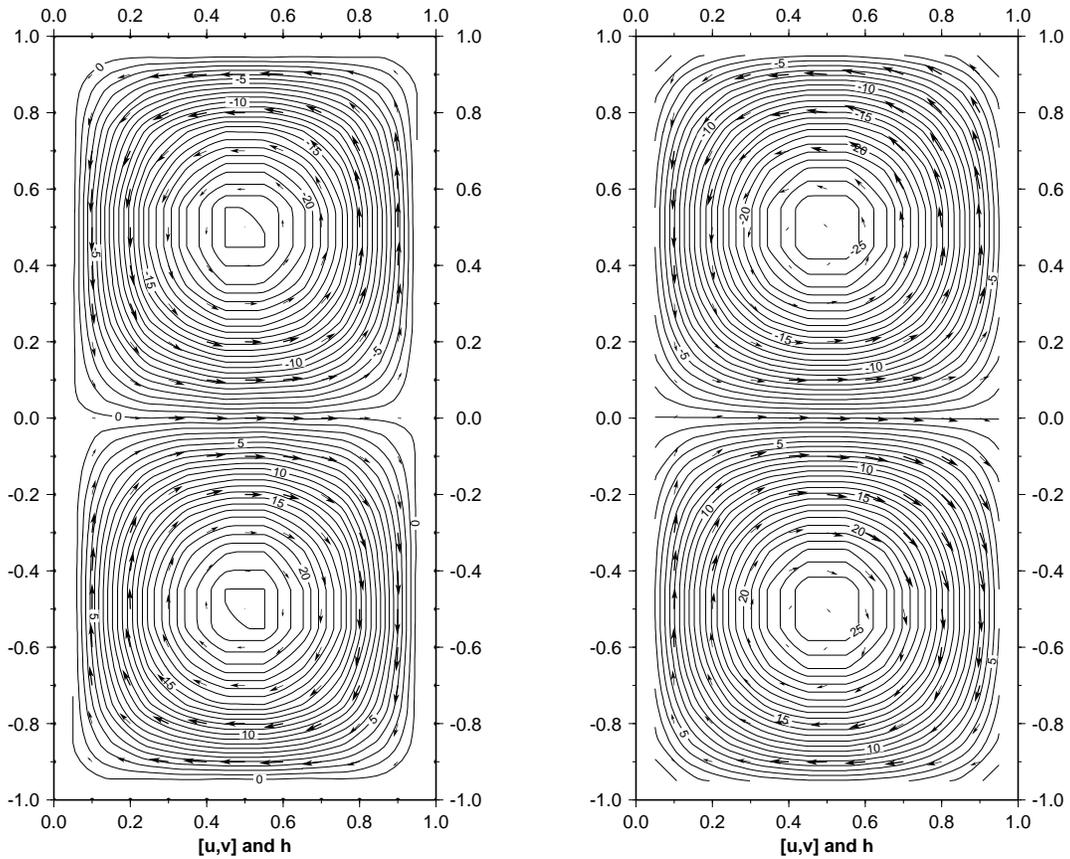


Figure 4.8: High resolution ($4S/\Delta x^2 = 5$) integrations of the B grid (left) and C grid (right) shallow water models. Shown are pressure (contours) and velocities (arrows) at time $t = 100$. Contour interval for pressure is 1. The models are approaching the steady state so the flow is nearly geostrophic. But for slightly different maximum values, the models would appear to be in good agreement. The divergence patterns (next figure) show that this is not the case.

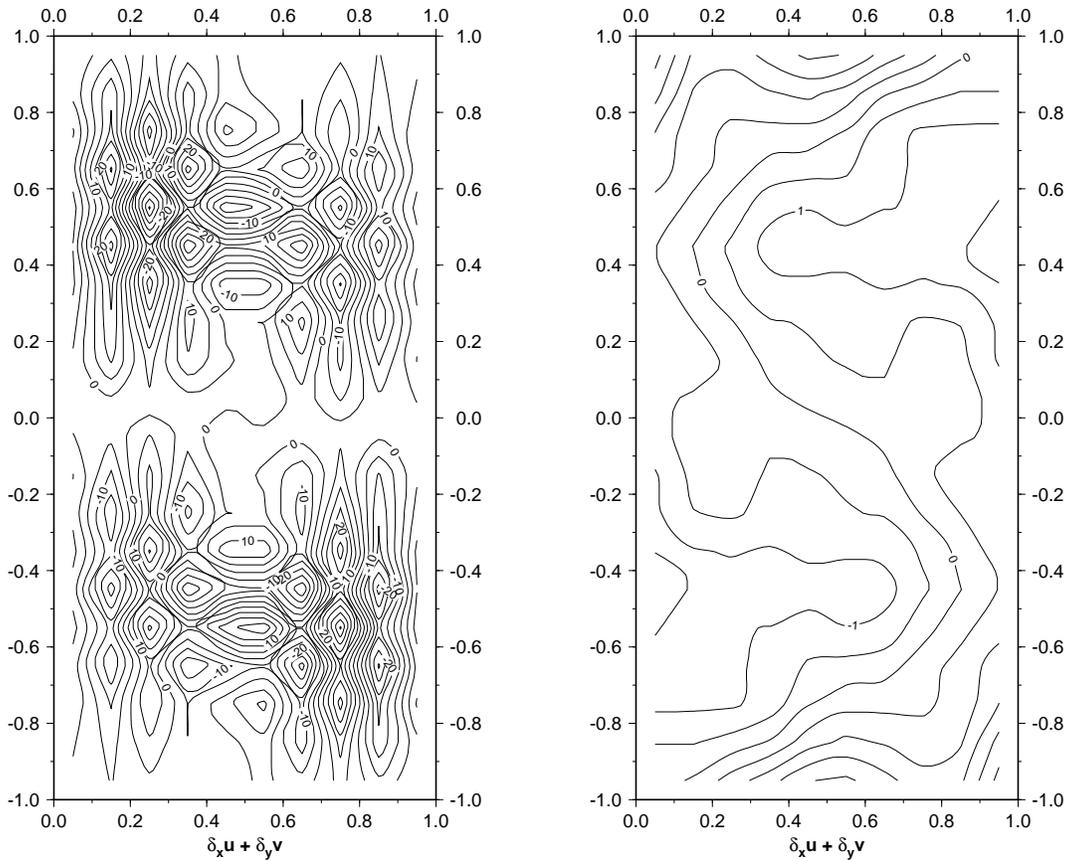


Figure 4.9: High resolution ($4S/\Delta x^2 = 5$) integrations of the B grid (left) and C grid (right) shallow water models. Shown is the horizontal divergence at time $t = 100$. Contour interval is 0.5 for the B grid and 5 for the C grid. The flow is nearly geostrophic at this stage in the integration so the divergence is small. The C grid clearly has no grid-scale noise. The B grid has a grid-scale signal, one order of magnitude larger than the smooth signal of the C grid.

4.6 The Explicit Oscillator Algorithm

In both B and C grid formulations, problems arise through the need to interpolate variables to their ‘unnatural’ positions. A $2\Delta x$ wave is invisible when viewed through the averaging operator $\overline{}$ implying that the physics represented by that term cannot act at that particular wave number. Thus on the B grid, a $2\Delta x$ wave cannot be aware of the presence of pressure gradient forces normal to the wave vector, though this is relatively unimportant at large scales. Similarly on the C grid, a $2\Delta x$ wave cannot feel Coriolis effects. In short, averaging is an undesirable feature in both these finite difference models.

In the case of the C grid, examination of equation (4.26) suggests that if one could ensure that spatial averaging did not occur in the oscillation term, $f^2\overline{D}^{tt}$, then the dispersion relation would approach the analytic limit to the maximum extent possible using second order finite differencing. This can be achieved as follows. Consider the forced oscillator form of the horizontal momentum equations:

$$\left. \begin{array}{l} u = u_o \\ v = v_o \end{array} \right\} \quad t = 0 \quad (4.33a)$$

$$\left. \begin{array}{l} \frac{\partial}{\partial t}u - fv = X \\ \frac{\partial}{\partial t}v + fu = Y \end{array} \right\} \quad t \geq 0 \quad (4.33b)$$

Here, $X = X(x, y, z, t)$ and $Y = Y(x, y, z, t)$ contain all other terms in the horizontal momentum equations. This is completely equivalent to the system:

$$\left. \begin{array}{l} u = u_o \\ v = v_o \\ \frac{\partial}{\partial t}u - fv = X \\ \frac{\partial}{\partial t}v + fu = Y \end{array} \right\} \quad t = 0 \quad (4.34a)$$

$$\left. \begin{array}{l} \frac{\partial^2}{\partial t^2}u - f\frac{\partial}{\partial t}v = \frac{\partial}{\partial t}X \\ \frac{\partial^2}{\partial t^2}v + f\frac{\partial}{\partial t}u = \frac{\partial}{\partial t}Y \end{array} \right\} \quad t \geq 0 \quad (4.34b)$$

By substituting for the acceleration $f\frac{\partial}{\partial t}(u, v)$ in equations 4.34b from 4.33b, the complete system can be written as a set of four initial conditions and two oscillator equations coupled through the forcing functions X and Y :

$$\left. \begin{array}{l} u = u_o \\ v = v_o \\ \frac{\partial}{\partial t}u - fv = X \\ \frac{\partial}{\partial t}v + fu = Y \end{array} \right\} \quad t = 0 \quad (4.35a)$$

$$\left. \begin{array}{l} \frac{\partial^2}{\partial t^2}u + f^2u = \frac{\partial}{\partial t}X + fY \\ \frac{\partial^2}{\partial t^2}v + f^2v = \frac{\partial}{\partial t}Y - fX \end{array} \right\} \quad t \geq 0 \quad (4.35b)$$

A discrete form for 4.35b might take the form:

$$\delta_{tt}u + f^2\overline{u}^{tt} = \delta_t X + f\overline{Y}^{txy} \quad (4.36a)$$

$$\delta_{tt}v + f^2\overline{v}^{tt} = \delta_t Y - f\overline{X}^{txy} \quad (4.36b)$$

and so propagation of the inviscid inertia-gravity waves would be determined by:

$$\frac{1}{f^2}\delta_{tt}D + \overline{D}^{tt} - L_\rho^2(\delta_{xx}D + \delta_{yy}D) = 0 \quad (4.37)$$

which leads to the dispersion relation:

$$\frac{4}{f^2\Delta t^2} \sin^2 \frac{\omega\Delta t}{2} = \cos^2 \frac{\omega\Delta t}{2} + \frac{4L_\rho^2}{\Delta x^2} \left(\sin^2 \frac{k\Delta x}{2} + \sin^2 \frac{l\Delta x}{2} \right) \quad (4.38)$$

Unlike their counter-parts on a C grid, equation 4.37 does not have any averaging operators at all. This is because the pressure terms within the interpolated operators of 4.36 cancel and so do not feature in 4.37. For this reason, it can be expected that the dispersion relation will not suffer from the deficiencies of the C grid formulation.

Assuming a continuous time derivative, the amplifying matrix method for finite difference equations can be applied to the formulation (4.36) of the inviscid shallow water equations (for which $X = -g'\delta_x h$, $Y = -g'\delta_y h$):

$$\begin{pmatrix} -\omega^2 + f^2 & 0 & \frac{2g'}{\Delta x}(\omega s_k - \imath f s_l c_k c_l) \\ 0 & -\omega^2 + f^2 & \frac{2g'}{\Delta x}(\omega s_l + \imath f s_k c_k c_l) \\ \imath \frac{2H}{\Delta x} s_k & \imath \frac{2H}{\Delta x} s_l & -\imath \omega \end{pmatrix} \begin{pmatrix} u_o \\ v_o \\ h_o \end{pmatrix} e^{\imath(kx+ly-\omega t)} = 0 \quad (4.39)$$

Setting the determinant of the amplification matrix to zero now yields five modes:

$$\omega = 0 \quad (4.40a)$$

$$\omega = \pm f \quad (4.40b)$$

$$\omega = \pm f \sqrt{1 + \frac{4L_\rho^2}{\Delta x^2} \left(\sin^2 \frac{k\Delta x}{2} + \sin^2 \frac{l\Delta x}{2} \right)} \quad (4.40c)$$

The new pair of modes ($\omega = \pm f$) is a consequence of manipulating the momentum equations into the form (4.35b). A free mode of arbitrary amplitude satisfies both sides of the equation, unless X and Y have some asymmetric properties. However, it should be noted that the dispersion relation for the inertia-gravity waves now is as close to the analytical relation as is possible with second order finite differencing. This has only been achieved before by Eliassen [MA76] using the D-D double grid scheme.

It would seem that the above integration scheme has advantages over the standard procedure on a C grid.

- The $2\Delta x$ wave can now feel the rotating frame and so cannot be excited as a stationary wave.
- The group speed is now of the right sign for all wave numbers (a property the B grid lacks).
- Elimination of the interpolated Coriolis terms permits a very simple implementation of a fully implicit coriolis term at no computational expense (on the C grid, this would normally require iteration of some sort - see Xu 1994 [Xu94]).

The potential advantages offered by the integration procedure certainly warrant investigation but as it stands, the existence of the extra mode pair potentially prohibits application of the scheme. A method that to all intents and purposes eliminates the mode is described next.

4.6.1 Implicit Coriolis: damping of the Extra Inertial Mode

The simplest treatment for the extra mode exploits the ability to now treat the Coriolis terms implicitly. Instead of equations 4.36, the finite difference scheme is:

$$\begin{aligned} \frac{(1 + \Delta t^2 f^2)u^{n+1} - 2u^n + u^{n-1}}{\Delta t^2} &= \frac{X^{n+\frac{1}{2}} - X^{n-\frac{1}{2}}}{\Delta t} + \overline{fY^{n+\frac{1}{2}}}^{xy} \\ \frac{(1 + \Delta t^2 f^2)v^{n+1} - 2v^n + v^{n-1}}{\Delta t^2} &= \frac{Y^{n+\frac{1}{2}} - Y^{n-\frac{1}{2}}}{\Delta t} - \overline{fX^{n+\frac{1}{2}}}^{xy} \end{aligned} \quad (4.41)$$

The time dependent part of the inertia-gravity waves is now damped according to $e^{-\imath\omega\Delta t} = e^{-\imath\omega_{Re}\Delta t} e^{\omega_{Im}\Delta t}$ where $\omega = \omega_{Re} + \imath\omega_{Im}$ are given by:

$$(1 + \Delta t^2 f^2)e^{-\imath\omega\Delta t} = 1 \pm \imath f \Delta t \quad \text{or} \quad \begin{aligned} \tan \omega_{Re} \Delta t &= \pm f \Delta t \\ \omega_{Im} \Delta t &= -\ln |1 + \Delta t^2 f^2| \end{aligned} \quad (4.42)$$

The inertial oscillations have a modified frequency but are damped for all $\Delta t > 0$ which makes the scheme unconditionally stable (thus allowing large time-steps). The natural frequencies of the shallow water system are now given by:

$$\omega = 0 \quad (4.43a)$$

$$\frac{2}{\Delta t} \sin \frac{\omega \Delta t}{2} = \pm f + \imath \lambda \quad (4.43b)$$

$$\frac{2}{\Delta t} \sin \frac{\omega \Delta t}{2} = \pm f \sqrt{1 + \frac{4L_\rho^2}{\Delta x^2} \left(\sin^2 \frac{k\Delta x}{2} + \sin^2 \frac{l\Delta x}{2} \right)} + \imath \lambda \quad (4.43c)$$

where λ is the amplitude decay factor per time step.

4.6.2 Comparison to the standard models

Figures 4.10 and 4.11 show the circulation, pressure and divergence fields at $t = 100$, for a C grid model that evaluates the momentum equations in the manner described above. The external parameters were identical to the analogous B and pure C grid integrations. A consequence of implementing the method just described is that the damping rates due to the implicit evaluation of Coriolis and the Adams-Bashforth extrapolation of the previous two models are different. This should not affect the final steady state but may change the convergence rate and phase of transients.

We conclude from these integrations that this ‘oscillator’ scheme does effectively correct the source of grid-scale noise resulting from inertia-gravity waves on the C grid at low resolution. At high resolution, the solution is unaffected (ie. is as accurate as the pure C grid) though the phase of the solution will obviously only match that of an implicit C grid model.

The scheme is, however, flawed for a physical reason which has not yet been discussed. Figure 4.12 shows the double gyre circulation on a β -plane obtained using the above scheme. A two-grid length wave, parallel to the meridional boundaries is clearly evident and seems to emanate from the western boundary.

A brief explanation of its origins is as follows. In the long term, the flow should become geostrophically balanced. Normally, on a ‘C’ grid, geostrophic balance takes the form:

$$\overline{f u^{xy}} = -\delta_y h \quad \overline{f v^{yx}} = \delta_x h \quad (4.44)$$

Now, the steady state form of equation 4.36 defines the geostrophic flow as:

$$f u = -\delta_y \overline{h^{xy}} \quad f v = \delta_x \overline{h^{xy}} \quad (4.45)$$

The averaging operator now appears in the slowly evolving geostrophic component of the flow. Grid-scale noise in the pressure field is filtered by the interpolation and not felt by the geostrophic circulation. Indeed, the solution (figure 4.12) shows a perfectly smooth Sverdrupian circulation, despite the high noise level in the pressure and divergence fields.

In summary, just as the Coriolis term on a C grid is invisible to grid scale noise in the flow field, the pressure gradient term in the geostrophic balance is invisible to grid scale noise in the pressure field. This implies that the shortest Rossby waves are poorly represented.

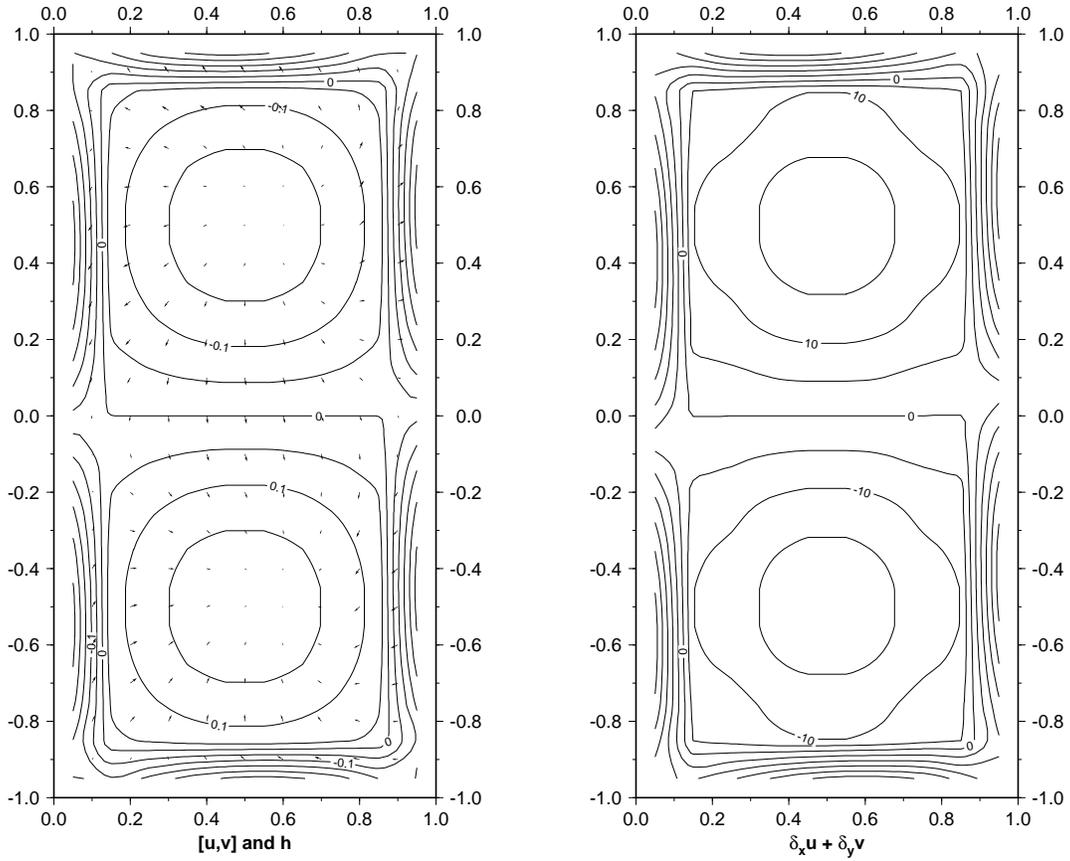


Figure 4.10: Low resolution ($4S/\Delta x^2 = 1/5$) integrations of the explicit oscillator scheme on the C grid. The tick marks indicate the grid-spacing. Shown are the circulation and pressure fields (left) and the horizontal divergence (right) at time $t = 100$. Contour intervals are 5 and 1 for the pressure and divergence respectively. All aspects of the fields are similar to those of the B grid integrations. The grid-scale noise normally associated with the C grid at this resolution is no longer apparent.

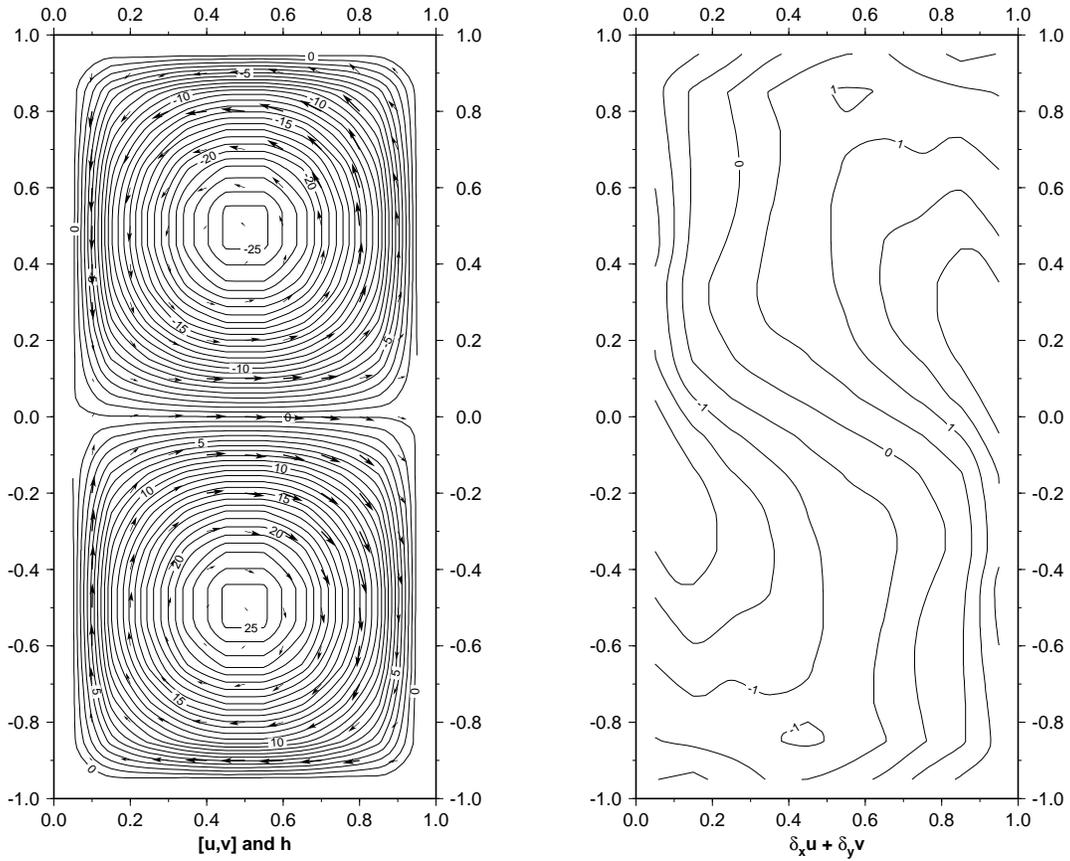


Figure 4.11: High resolution ($4S/\Delta x^2 = 5$) integrations of the explicit oscillator scheme on the C grid. The tick marks indicate the grid-spacing. Shown are the circulation and pressure fields (left) and the horizontal divergence (right) at time $t = 100$. Contour intervals are 5 and 1 for the pressure and divergence respectively. The divergence field is as smooth as, though out of phase with, the divergence pattern of the pure C grid integration.

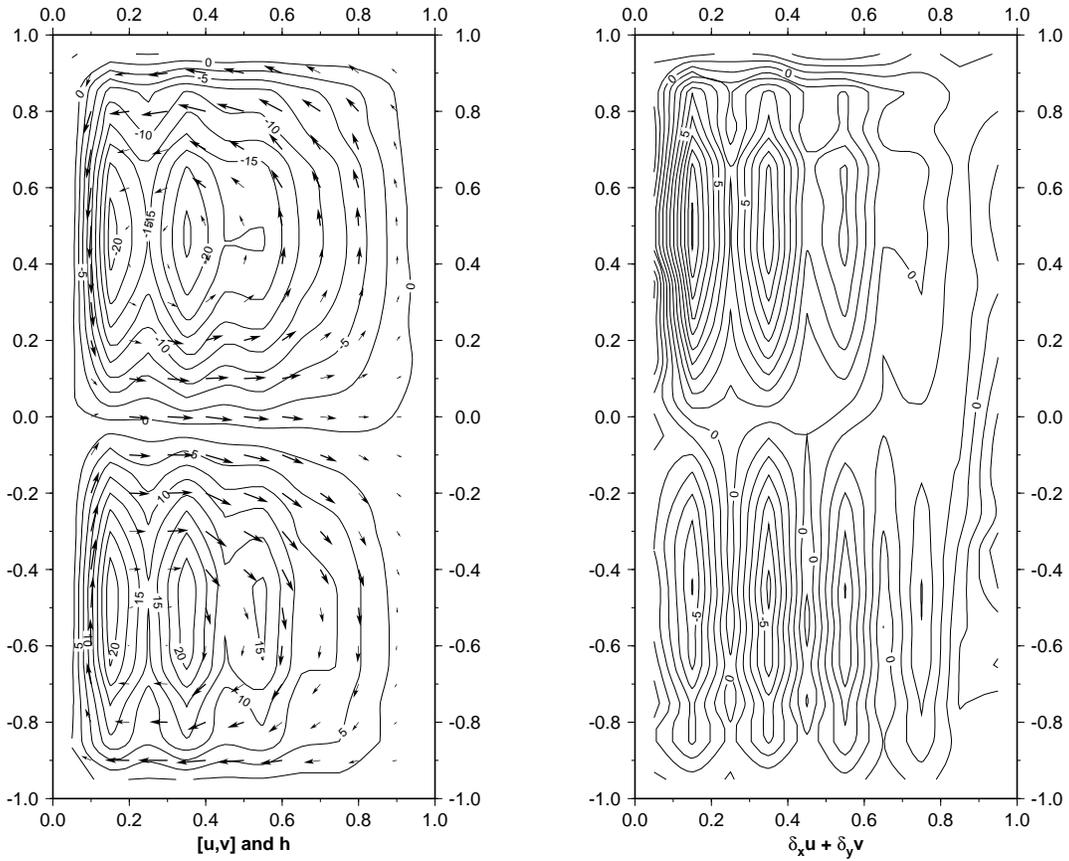


Figure 4.12: Explicit Oscillator scheme at high resolution ($4S/\Delta x^2 = 1/2$) integrations on a β -plane. Shown are the circulation and pressure fields (left) and the horizontal divergence (right) at time $t = 100$. Contour intervals are 2.5 and 1 for the pressure and divergence respectively. Despite the good performance of inertia-gravity waves, a two-grid length wave, parallel to the meridional boundaries, emanating from the western boundary is prevalent.

4.7 Summary

The GCM suffered from grid-scale noise when the Rossby radius of deformation was not resolved. An analysis of the behavior of inertia-gravity waves in numerical models led to an understanding of the origin of the noise.

An alternative “oscillator” formulation of the momentum equations was proposed for implementation on the C grid. When compared with the solutions from the B and normal C grids, the scheme appears to successfully represent the short inertia-gravity waves on the f plane. However, in the presence of a planetary vorticity gradient, the approach fails to represent Rossby waves adequately.

Chapter 5

Numerical Representation of Rossby Waves and the C_d scheme

Using the methods that were employed to understand inertia-gravity waves on numerical grids, the behaviour of Rossby waves can also be analysed. Rossby waves are a consequence of the Taylor-Proudman theorem and the geometry of the earth¹. Motions of sufficiently large scale can “feel” the variation in the Coriolis parameter. Because of the longer spatial scales, the time scales involved are typically longer than the rotation period of the earth.

The following section develops shallow water theory further to study the Rossby wave mode of motion. The equivalent discrete analysis is then applied to grids studied in the previous chapter. Inspired by the “explicit oscillator” scheme developed for inertia-gravity waves, a new approach (the C_d scheme) is described that also treats Rossby waves with fidelity. The new scheme is analysed, tested and compared to the standard models. Finally, it is implemented in the GCM and its performance evaluated.

5.1 Rossby Waves

In chapter 4, the shallow water equations were derived assuming that the horizontal scale of motion was small enough that variations in the Coriolis parameter could be assumed negligible, $\beta \ll 1$. On the basin scale, this assumption must be relaxed. Further, the time-scale associated with these large spatial scales is much longer than the rotation period, $\frac{1}{f_o T} \ll 1$.

Allowing the variation in Coriolis parameter introduces Rossby waves into the model. The flow to the first two orders in some arbitrary small parameter, ϵ , can be written:

$$u = u_o + \epsilon u_1 + O(\epsilon^2) \quad (5.1)$$

Substituting into the shallow water equations 4.12 and keeping only the order one terms yields:

$$\hat{\mathbf{k}} \wedge \mathbf{u}_o + \nabla h = 0 \quad (5.2a)$$

$$S \nabla \cdot \mathbf{u}_o = 0 \quad (5.2b)$$

which are self consistent equations describing a quasi-geostrophic flow. The difference between a quasi-geostrophic model and a geostrophic model is that a reference Coriolis parameter is used in quasi-geostrophy, rather than the local value, in defining the geostrophic flow.

Defining the zeroth order flow by 5.2 and then substituting back into the shallow water equations and retaining all terms of order $\frac{1}{f_o T}$, R_o , β and ϵ yields:

$$\frac{1}{f_o T} \frac{\partial}{\partial t} \mathbf{u}_o + \epsilon \hat{\mathbf{k}} \wedge \mathbf{u}_1 + \beta y \hat{\mathbf{k}} \wedge \mathbf{u}_o = 0 \quad (5.3a)$$

$$\frac{1}{f_o T} \frac{\partial}{\partial t} h + \epsilon S \nabla \cdot U_1 = 0 \quad (5.3b)$$

which describe the slow evolution of the quasi-geostrophic flow in the presence of β . The ageostrophic component of velocity can be eliminated by taking the curl of the momentum equation to form the vorticity equation and

¹Strictly speaking, Rossby waves propagate in a Potential Vorticity gradient which does not necessarily require β .

substituting from the continuity equation:

$$\frac{1}{f_o T} \frac{\partial}{\partial t} \zeta_o + \beta v_o - \frac{1}{f_o T} \frac{1}{S} \frac{\partial}{\partial t} h = 0 \quad (5.4)$$

The vorticity of the quasi-geostrophic flow can be written $\zeta_o = \nabla^2 h$ and so equation 5.4 can be written in terms of one variable, yielding the potential vorticity equation:

$$\frac{1}{f_o T} \frac{\partial}{\partial t} \nabla^2 h + \beta \frac{\partial}{\partial x} h - \frac{1}{f_o T} \frac{1}{S} \frac{\partial}{\partial t} h = 0 \quad (5.5)$$

The picture of this mode then, is one of a flow field in quasi-geostrophic equilibrium with the pressure gradients (the zeroth order balance), and a slowly evolving pressure field (the first order balance). The dispersion relation for linear plane Rossby waves is obtained directly from 5.5:

$$\omega = \frac{-\beta S k}{1 + S(k^2 + l^2)} \quad (5.6)$$

Notice that the denominator takes the form of the square of the inertia-gravity wave frequency. The non-dimensional² frequency is plotted as the first surface in figures 5.1 and 5.2. The β effect is the cause of the anisotropic propagation.

The phase velocity:

$$\begin{pmatrix} \omega/k \\ \omega/l \end{pmatrix} = \frac{\beta S}{1 + S(k^2 + l^2)} \begin{pmatrix} -1 \\ -k/l \end{pmatrix} \quad (5.7)$$

is always pointed towards the west with a maximum westward velocity of $\epsilon \beta S$ (in dimensional units $c_{max}^\phi = \beta L_D^2$). The phase velocity behaves as $1/f^2$ being larger near the equator than at the poles.

The group velocity:

$$\begin{pmatrix} \frac{\partial \omega}{\partial k} \\ \frac{\partial \omega}{\partial l} \end{pmatrix} = \frac{\beta S}{(1 + S(k^2 + l^2))^2} \begin{pmatrix} k^2 - l^2 - \frac{1}{S} \\ \frac{2kl}{S} \end{pmatrix} \quad (5.8)$$

can be directed either eastward or westward depending on the sign of $k^2 - l^2 - 1/S$. For short zonal waves $k^2 > l^2 + 1/S$, the group velocity has an eastward component whilst for long zonal waves $k^2 < l^2 + 1$, the zonal group velocity is westward. Thus, whilst all wave fronts propagate westward, only long waves transmit energy westward. Short waves transmit energy eastward. This anisotropy is one interpretation of the phenomena of western boundary intensification [Ped79].

The turning point in zonal group speed ($\frac{\partial \omega}{\partial k} = 0$) occurs at $S k^2 = 1 + S l^2$. The maximum frequency thus occurs at $l = 0$, $k = 1/\sqrt{S}$ with a frequency $\beta\sqrt{S}/2$. In dimensional units this is a maximum frequency of $\beta L_D/2$ and is typically of the order of a few cycles per month.

5.2 Finite-differenced Rossby waves

The quasi-geostrophic scaling applied in the previous section to obtain the Rossby wave mode of motion can be applied to the finite difference models. The analysis (see Fox-Rabinovitz [FR91] and Wajsowicz [Waj86]) is summarised as follows.

For the A grid:

$$\text{Geostrophic: } v_g = \delta_x \bar{h}^x \quad u_g = -\delta_y \bar{h}^y \quad (5.9a)$$

$$\text{PV equation: } \frac{\partial}{\partial t} (\delta_{xx} \bar{h}^{xx} + \delta_{yy} \bar{h}^{yy} - \frac{1}{S} h) + \beta (\delta_x \bar{h}^{xyy} - \frac{\Delta y^2}{4} \delta_{xyy} \bar{h}^x) = 0 \quad (5.9b)$$

$$\text{Dispersion: } \omega = \frac{\epsilon \beta S s_k c_k (c_l^2 + \Delta y^2 s_l^2 / 4)}{1 + S(s_k^2 c_k^2 + s_l^2 c_l^2)} \quad (5.9c)$$

For the B grid:

$$\text{Geostrophic: } v_g = \delta_x \bar{h}^y \quad u_g = -\delta_y \bar{h}^x \quad (5.10a)$$

$$\text{PV equation: } \frac{\partial}{\partial t} (\delta_{xx} \bar{h}^{yy} + \delta_{yy} \bar{h}^{xx} - \frac{1}{S} h) + \beta \delta_x \bar{h}^x = 0 \quad (5.10b)$$

$$\text{Dispersion: } \omega = \frac{\epsilon \beta S s_k c_k}{1 + S(s_k^2 c_l^2 + s_l^2 c_k^2)} \quad (5.10c)$$

²Time was non-dimensionalised with respect to f in chapter 4.

For the C grid:

$$\text{Geostrophic: } \overline{v}_g^{xy} = \delta_x h \quad \overline{u}_g^{xy} = -\delta_y h \quad (5.11a)$$

$$\text{PV equation: } \frac{\partial}{\partial t}(\delta_{xx}h + \delta_{yy}h - \frac{1}{S}\overline{h}^{xyy}) + \beta\delta_x\overline{h}^{xyy} = 0 \quad (5.11b)$$

$$\text{Dispersion: } \omega = \frac{\epsilon_\beta S s_k c_k c_l^2}{c_k^2 c_l^2 + S(s_k^2 + s_l^2)} \quad (5.11c)$$

For the D grid:

$$\text{Geostrophic: } v_g = \delta_x h \quad u_g = -\delta_y h \quad (5.12a)$$

$$\text{PV equation: } \frac{\partial}{\partial t}(\delta_{xx}h + \delta_{yy}h - \frac{1}{S}h) + \beta\delta_x\overline{h}^{xyy} = 0 \quad (5.12b)$$

$$\text{Dispersion: } \omega = \frac{\epsilon_\beta S s_k c_k c_l^2}{1 + S(s_k^2 + s_l^2)} \quad (5.12c)$$

For the explicit oscillator scheme on a C grid:

$$\text{Geostrophic: } v_g = \delta_x \overline{h}^{xy} \quad u_g = -\delta_y \overline{h}^{xy} \quad (5.13a)$$

$$\text{PV equation: } \frac{\partial}{\partial t}(\delta_{xx}h + \delta_{yy}h - \frac{1}{S}h) + \beta\delta_x\overline{h}^{xyy} = 0 \quad (5.13b)$$

$$\text{Dispersion: } \omega = \frac{\epsilon_\beta S s_k c_k c_l^2}{1 + S(s_k^2 + s_l^2)} \quad (5.13c)$$

It is interesting to note that the explicit oscillator scheme, despite a very different representation of the physics, exhibits the same Rossby wave dispersion as the D grid.

One observation about the dispersion relations, as written, is that the denominator of the expressions for the frequency take the form of the dispersion relation of the inertia-gravity waves for the respective grid. Therefore, the denominator is likely to have some artificial zero points creating singularities in the frequency. In fact, this does not become a problem since wherever a singularity might occur, the numerator has an artificial zero that vanishes more rapidly.

Figures 5.1 and 5.2 show the discrete dispersion relations plotted for high and low wave resolutions respectively. Included in the the diagram is the dispersion relation pertaining to the Eliassen time-staggered grid that will be described later in this chapter.

Both the high and low resolution dispersion relation for the A grid have artificial turning points. Simply put, the shorter A grid Rossby waves bear no resemblance in dispersion properties to the continuum. Of the other high resolution dispersion relations (figure 5.1) only the B grid exhibits a qualitative difference to the continuum. This takes the form of artificially fast zonal group speed for long meridional waves. The B, C and D grids have zero frequencies for zonal grid length waves but since the continuum has very low frequencies here, the qualitative behaviour is unchanged.

As was the case for the inertia-gravity waves, the dispersion relations for the different grids dramatically change their behaviour as one moves from high to low resolution. Figure 5.2 shows the frequency for short zonal waves. None of the grids reproduce the behaviour of the continuum plot properly.

All the models except the B grid have artificially high meridional group speeds. This means that the models are able to meridionally redistribute energy in an artificial manner.

One point in favour of the C grid at low resolution is that for short meridional waves, the zonal group speed is much closer to the continuum than the B grid (*i.e.* the artificial turning point is at a shorter wave number on the C grid than on the B grid). Indeed, the B grid has the wrong sign for several of the shortest zonal wave numbers whilst the C grid is in error for only the shortest. This has implications for western boundary intensification in the models. If waves that are longer than the grid-scale can transmit energy eastward then the western boundary layer may be artificially wider as a result.

It is interesting to note that for the D grid, despite a very poor representation of inertia-gravity waves, the Rossby waves seem to be well represented. This is due to the natural geostrophic balance that has no spatial averaging in it.

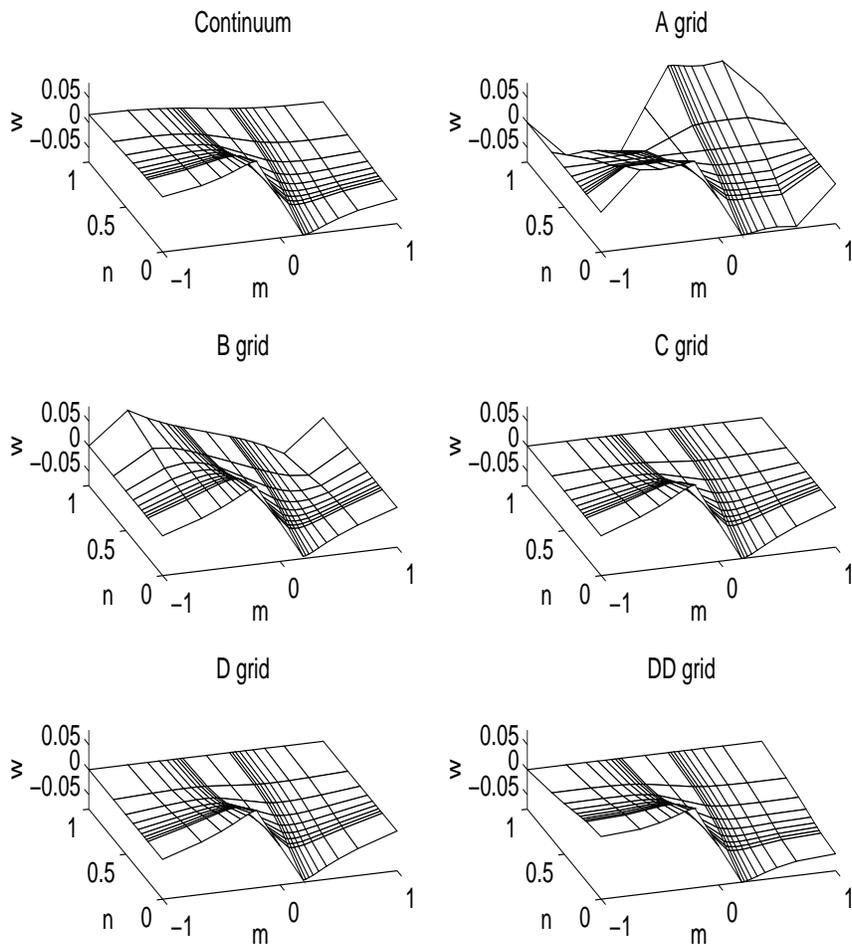


Figure 5.1: Rossby wave dispersion relations for the continuum and high resolution finite difference models (A,B,C,D,DD) with a wave resolution of $2\sqrt{S}/\Delta x = 3$.

5.3 Time staggered grids

It should be clear from the results of using the explicit oscillator scheme that both classes of discrete wave motion need to be taken into account; mis-representation of the inertia-gravity waves generates erroneous divergence patterns on a short time scale whilst mis-representation of Rossby waves generates spurious standing waves on a much longer time scale.

Of the gridding schemes discussed so far, none of them satisfactorily meets both requirements at all wave resolutions. To devise a scheme that can achieve this goal, one might begin by considering one that performs well at just one scale. Here, the C grid performs particularly well at high wave resolution and the B grid is perhaps to be favoured at low resolution (this decision is based upon the accurate meridional propagation of energy rather than the scale of the western boundary current). However, since the non-hydrostatic physics in the GCM is most naturally described on a C grid, the following discussion will concentrate upon the C grid structure.

There are in fact many gridding schemes that can be introduced when more than one time step is used in the time-stepping algorithm. This involves staggering variables not only in space but also in time. This was already an implicit feature of the single time step schemes of the previous B and C grid models. The pressure variable was staggered in time with the velocity variables allowing second order finite differencing in both time and space for the gravity waves.

Five gridding schemes derived from the five Arakawa grids were described by Eliassen [MA76, FR91]. The method shifts the grid diagonally at alternating time-steps. Figure 5.3 shows the three dimensional distribution of variables in space and time for the five grids. The primed notation indicates that the grid is shifted diagonally in space from the unprimed position. The models are described as follows:

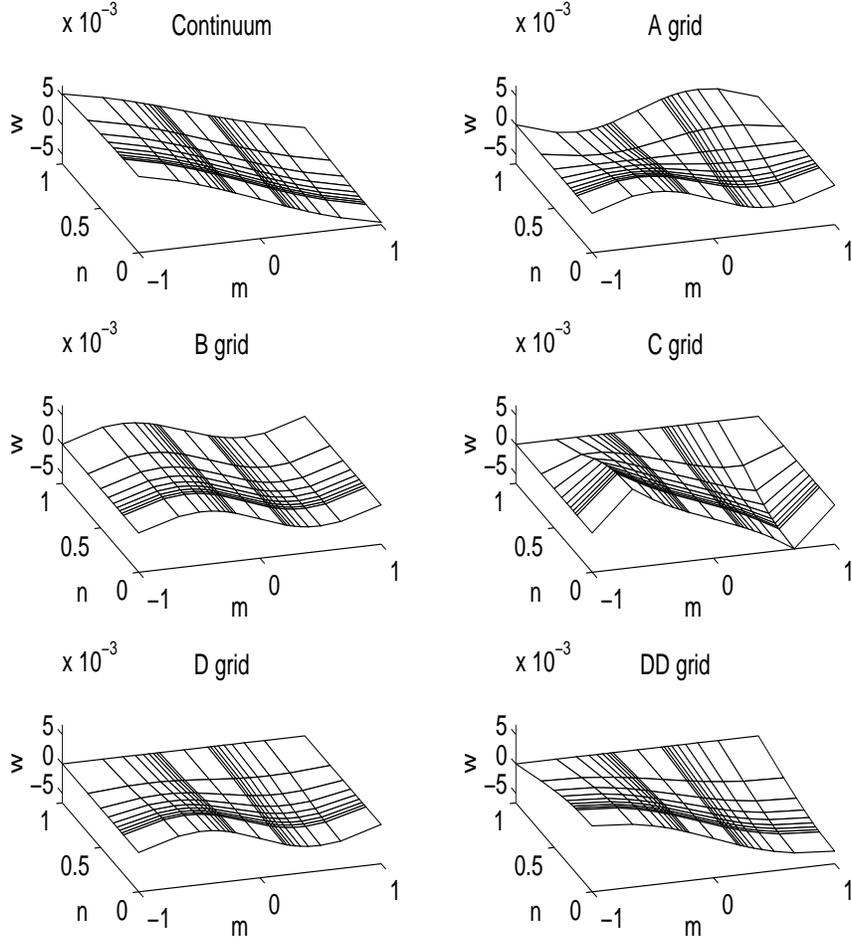


Figure 5.2: Rossby wave dispersion relations for the continuum and low resolution finite difference models (A,B,C,D,DD) with a wave resolution of $2\sqrt{S/\Delta x} = 1/3$.

Scheme AA':

$$\delta_t \bar{u}^t - \bar{v}^{xy} + \delta_x \bar{h}^y = 0 \quad (5.14a)$$

$$\delta_t \bar{v}^t + \bar{u}^{xy} + \delta_y \bar{h}^x = 0 \quad (5.14b)$$

$$\delta_t \bar{h}^t + S(\delta_x \bar{u}^y + \delta_y \bar{v}^x) = 0 \quad (5.14c)$$

$$\text{IG dispersion: } \omega^2 = c_k^2 c_l^2 + S(s_k^2 c_l^2 + s_l^2 c_k^2) \quad (5.14d)$$

Scheme BB':

$$\delta_t \bar{u}^t - \bar{v}^{xy} + \delta_x \bar{h}^x = 0 \quad (5.15a)$$

$$\delta_t \bar{v}^t + \bar{u}^{xy} + \delta_y \bar{h}^y = 0 \quad (5.15b)$$

$$\delta_t \bar{h}^t + S(\delta_x \bar{u}^x + \delta_y \bar{v}^y) = 0 \quad (5.15c)$$

$$\text{IG dispersion: } \omega^2 = c_k^2 c_l^2 + S(s_k^2 c_k^2 + s_l^2 c_l^2) \quad (5.15d)$$

Scheme CC':

$$\delta_t \bar{u}^t - v + \delta_x \bar{h}^{xy} = 0 \quad (5.16a)$$

$$\delta_t \bar{v}^t + u + \delta_y \bar{h}^{xy} = 0 \quad (5.16b)$$

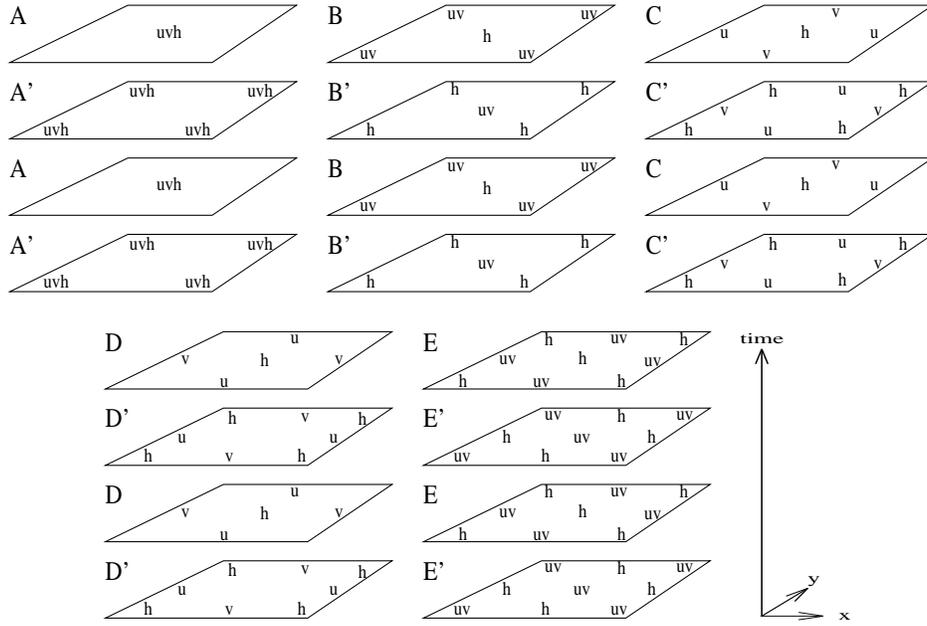


Figure 5.3: The Eliassen time horizontally staggered grids, AA',BB',CC',DD',EE'

$$\delta_t \bar{h}^t + S(\delta_x \bar{u}^{xy} + \delta_y \bar{v}^{xy}) = 0 \quad (5.16c)$$

$$\text{IG dispersion: } \omega^2 = 1 + S c_k^2 c_l^2 (s_k^2 + s_l^2) \quad (5.16d)$$

Scheme DD':

$$\delta_t \bar{u}^t - v + \delta_x h = 0 \quad (5.17a)$$

$$\delta_t \bar{v}^t + u + \delta_y h = 0 \quad (5.17b)$$

$$\delta_t \bar{h}^t + S(\delta_x u + \delta_y v) = 0 \quad (5.17c)$$

$$\text{IG dispersion: } \omega^2 = 1 + S(s_k^2 + s_l^2) \quad (5.17d)$$

Again, the EE' grid is a rotated and scaled (by $\sqrt{2}$) version of the BB' grid and so omitted. All the above models use explicit evaluation of both Coriolis and pressure terms. The spatial interpolation of these terms can actually be reduced or even avoided by semi-implicit (centered over two time levels) evaluation. For example the Coriolis terms of the AA' grid could replace the $-xy$ with a $-2t$ with no loss of efficiency or accuracy. Here, only the fully explicit models will be considered to allow an easy transfer of algorithms to more sophisticated models.

Before continuing with a closer look at these time staggered grids, there are a number of cautionary points:

- The storage requirements of the computer system can be as much as doubled because there is a duplication of variables.
- There is a duplication of physical boundary conditions.
- There is a tendency for the two alternating grids to decouple (in the manner of a leap-frog computational mode); this can be alleviated by adding filtering terms.

The decoupling can be thought of as a consequence of the splitting of each of the natural modes of the system. So there might be two geostrophic modes and two pairs of inertia-gravity modes. The pair may look identical in dispersion qualities but the reciprocal pairs are linearly independent.

The grid that draws our attention is the DD' grid. No interpolation operators appear in the finite difference equations and thus the inertia-gravity wave dispersion relation is the same as that of the explicit oscillator scheme of the previous chapter.

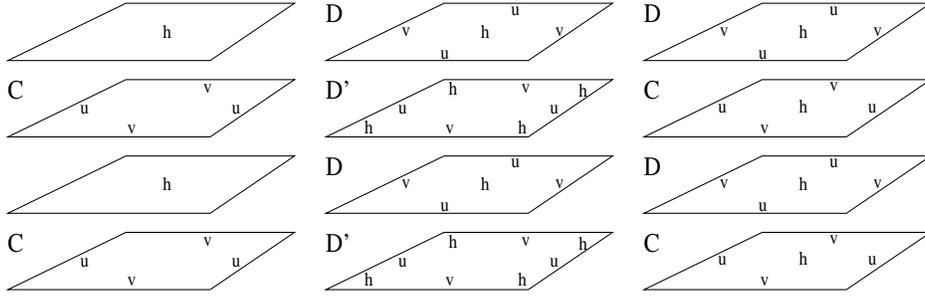


Figure 5.4: The single step C grid compared to the time staggered grids DD' CD

The DD' grid, despite its name, turns out to be based on the space-time structure of the normal C grid. Figure 5.4 shows the two grids side by side. It is clear that the DD' grid is just two C grid models superposed with the Coriolis terms evaluated using the more natural velocities. Thus the accurate gravity wave representation of the C grid is maintained and the pure inertial oscillation behaviour is corrected. The Rossby wave dispersion relation of the DD' grid is also an improvement upon the C grid:

$$\omega = \frac{\epsilon_\beta S s_k c_l}{1 + S(s_k^2 + s_l^2)} \quad (5.18)$$

and is plotted in the earlier figures 5.1 and 5.2.

The artificial zero frequency of zonal grid-scale waves, common to all the previous models at low resolution, is no longer apparent. This means that the scale of the western boundary will be of the right size or as small as the grid can resolve. Again at low resolution, the DD' grid still renders an artificially large meridional component of group velocity, not present on the B grid. In summary, the high resolution properties of the DD' grid are very similar to the C grid, and, overall, appears to be an improvement upon the C grid.

5.4 The C_d grid scheme: the CD hybrid

A further degree of freedom in devising a scheme would be to change the spatial grid for alternate time-steps rather than simply shift it the grid diagonally. For example, a B grid could be staggered in time with a C grid, with the pressure points either spatially aligned or diagonally shifted. The full range of permutations is beyond description. However, one such combination of grids, inspired by the explicit oscillator scheme, is the CD grid shown in figure 5.4.

Here, the pressure variable is at the same point in space for all time levels. The horizontal velocities switch positions in time. The finite difference equations now have to be written for each grid, producing six prognostic equations:

$$\delta_t \bar{u}_C^t - v_D + \delta_x h_D = 0 \quad (5.19a)$$

$$\delta_t \bar{v}_C^t + u_D + \delta_y h_D = 0 \quad (5.19b)$$

$$\delta_t \bar{h}_D^{-t} + S \nabla \cdot \mathbf{u}_C = 0 \quad (5.19c)$$

$$\delta_t \bar{u}_D^t - v_C + \overline{\delta_x h_C}^{xy} = 0 \quad (5.19d)$$

$$\delta_t \bar{v}_D^t + u_C + \overline{\delta_y h_C}^{xy} = 0 \quad (5.19e)$$

$$\delta_t \bar{h}_C^{-t} + S \nabla \cdot \bar{\mathbf{u}}_D^{xy} = 0 \quad (5.19f)$$

As was alluded to earlier in the discussion about decoupling, there are six natural modes in these staggered systems, a result of six prognostic equations. The four inertia-gravity modes in this CD grid are:

$$\omega^2 = 1 + S(s_k^2 + s_l^2) \quad (5.20a)$$

$$\omega^2 = 1 + S c_k^2 c_l^2 (s_k^2 + s_l^2) \quad (5.20b)$$

One pair of inertia-gravity modes is clearly preferable to the other. The first pair takes the form of the explicit oscillator and DD' grids. The second pair contains cosines ($c_k^2 c_l^2$) resulting from spatial averaging of the gravity term.

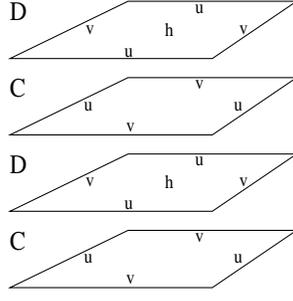


Figure 5.5: The two step C_d grid.

The issue of duplicated physics should be considered at this point. Not only is the number of modes doubled, but the duplicates (5.20b) are poor because they involve spatial averaging (signified by the cosines). A step towards reducing the number of natural modes, and towards decreasing the duplication of variables, is to eliminate the pressure variable at the C grid time level, as is shown in figure 5.5.

The system is reduced to five prognostic variables and equations:

$$\delta_t \bar{u}_C^t - v_D + \delta_x h = 0 \quad (5.21a)$$

$$\delta_t \bar{v}_C^t + u_D + \delta_y h = 0 \quad (5.21b)$$

$$\delta_t \bar{h}^t + S \nabla \cdot \mathbf{u}_C = 0 \quad (5.21c)$$

$$\delta_t \bar{u}_D^t - v_C + \overline{\delta_x h}^{xyt} = 0 \quad (5.21d)$$

$$\delta_t \bar{v}_D^t + u_C + \overline{\delta_y h}^{xyt} = 0 \quad (5.21e)$$

$$(5.21f)$$

with the five natural modes being:

$$\omega = 0 \quad (5.22a)$$

$$\omega^2 = 1 + S(s_k^2 + s_l^2) \quad (5.22b)$$

$$\omega^2 = 1 \quad (5.22c)$$

The remarkable aspect about these natural modes is that despite spatial interpolation operators in the prognostic equations, no such operators appear in the dispersion relations.

However, there remain two extraneous modes now taking the form of inertial oscillations. To filter out this pair of modes, a further reduction in prognostic variables would be necessary. Since this would recover one of the previous single time level models, the next best treatment is to damp the modes.

There are two ways to achieve this. The first is to over-extrapolate the Coriolis terms on the D grid to dampen the inertial oscillations.

$$\mathbf{u}_D^{n+\frac{1}{2}} = \mathbf{u}_D^{n-\frac{1}{2}} + \Delta t \left(-\hat{\mathbf{k}} \wedge \mathbf{u}_C^{n+\epsilon_{AB}} - \nabla \bar{h}^{xyt} \right) \quad (5.23)$$

where the $n + \epsilon_{AB}$ extrapolation takes the form $\mathbf{u}^{n+\epsilon_{AB}} = (1 + \epsilon_{AB})U^n - \epsilon_{AB}\mathbf{u}^{n-1}$.

It should be emphasized that this Adams-Bashforth extrapolation is *not* necessary for the C_d scheme to be stable. It is introduced solely for the purpose of the damping the inertial oscillations which are easily excited.

The second damping method is to tie or couple the D grid variables to the C grid variables over some finite period of time longer than one time step. Here, a coupling term is introduced into the D grid momentum equations of the form:

$$\frac{\partial}{\partial t} \mathbf{u}_D + \dots = -\hat{C}(\bar{\mathbf{u}}_D^t - \bar{\mathbf{u}}_C^{xy}) \quad (5.24)$$

where \hat{C} is some operator yet to be determined.

Without the coupling, the scheme has two artificial inertial modes and a geostrophic balance much like the explicit oscillator scheme ($\mathbf{u}_C = -\hat{\mathbf{k}} \wedge \nabla h$). The Rossby waves disperse according to:

$$\omega = \frac{-\epsilon_{\beta} S s_k c_k c_k^2}{1 + S(s_k^2 + s_l^2)} \quad (5.25)$$

which is the same as for the D grid and explicit oscillator scheme. In fact, on closer examination, the scheme is just the integral form of the second order equations employed in devising the explicit oscillator scheme.

If one fully relaxes the D grid velocities to the C grid in one time step, then they become equal, by definition, to the spatial average of the C grid velocities for all time. In this limit, the scheme becomes the normal single time level C grid.

For the case of coupling on a time scale longer than the inertial period but shorter than the period of the fastest Rossby wave ($\epsilon_\beta \sqrt{S}/2 < \lambda < 1$), the inertia-gravity waves will be relatively unaffected since the motion is on a much shorter time-scale than the relaxation. So the good short time scale performance of the explicit oscillator scheme is reproduced here. On time-scales longer than the relaxation, the D grid velocities are fully relaxed to the interpolated C grid values so the geostrophic balance takes the form of the C grid: $\bar{\mathbf{u}}^{xy} = -\hat{\mathbf{k}} \wedge \nabla h$.

This would yield the optimum attributes of the two inherent grids. It then remains to determine the form that the coupling operator should take.

5.5 Determining the optimal coupling

Damped inertia-gravity waves are described by the continuous equations:

$$\frac{\partial}{\partial t} \mathbf{u} + \hat{\mathbf{k}} \wedge \mathbf{u} + S \nabla h = -\hat{F}(\mathbf{u}) \quad (5.26a)$$

$$\frac{\partial}{\partial t} h + \nabla \cdot \mathbf{u} = -\hat{H}(h) \quad (5.26b)$$

where \hat{D} and \hat{H} are dissipative operators, yet to be specified. The corresponding divergence equation is:

$$\left[\left(\hat{F} + \frac{\partial}{\partial t} \right)^2 + 1 - S \left(\hat{H} + \frac{\partial}{\partial t} \right) \nabla^2 \right] \nabla \cdot \mathbf{u} = 0 \quad (5.27)$$

The finite difference equations using the C_d scheme, including the coupling term $\hat{C}(\mathbf{u}_D - \bar{\mathbf{u}}_C^{xy})$, are:

$$\frac{\partial}{\partial t} \mathbf{u}_C + \hat{\mathbf{k}} \wedge \mathbf{u}_D + S \nabla h = -\hat{F}(\mathbf{u}_C) \quad (5.28a)$$

$$\frac{\partial}{\partial t} \mathbf{u}_D + \hat{\mathbf{k}} \wedge \mathbf{u}_C + S \nabla \bar{h}^{xy} = -\hat{F}(\bar{\mathbf{u}}_C^{xy}) - \hat{C}(\mathbf{u}_D - \bar{\mathbf{u}}_C^{xy}) \quad (5.28b)$$

$$\frac{\partial}{\partial t} h + \nabla \cdot \mathbf{u}_C = -\hat{H}(h) \quad (5.28c)$$

where \hat{C} is still arbitrary.

The divergence ($\nabla \cdot \mathbf{u}_C$) equation, that governs the dispersion of the discrete inertia-gravity waves, is:

$$\left[\left(\hat{F} + \frac{\partial}{\partial t} \right) \left(\hat{C} + \frac{\partial}{\partial t} \right) + 1 - S \left(\hat{H} + \frac{\partial}{\partial t} \right) \nabla^2 \right] \nabla \cdot \mathbf{u}_C + (\hat{F} - \hat{C}) \bar{\zeta}_C^{xy} = 0 \quad (5.29)$$

which has a new term involving vorticity on the C grid. This term vanishes only if:

$$\hat{C} = \hat{F} \quad (5.30)$$

which indicates that the coupling should take the form of the dissipation in the momentum equations. The divergence wave equation then looks very much like the continuous system with no interpolation operators.

The natural modes of the system on an f-plane are given by:

$$\left(\left(\hat{F} + \frac{\partial}{\partial t} \right)^2 + 1 \right) \left\{ \left(\left(\hat{F} + \frac{\partial}{\partial t} \right)^2 + 1 \right) \left(\hat{H} + \frac{\partial}{\partial t} \right) - \left(\hat{F} + \frac{\partial}{\partial t} \right) S \nabla^2 \right\} h = 0 \quad (5.31)$$

The first factor gives rise to the inertial oscillations, now damped both by the explicit dissipation and that implied in the Adams-Bashforth time-stepping. The second factor corresponds to the damped inertia-gravity waves corresponding to 5.27. There are no spatial interpolation operators here, so the inertia-gravity waves should not have any artificial signs in phase and group speed.

	Experiment I	Experiment II
Burger number S	6.25×10^{-4}	6.25×10^{-4}
Planetary β	0.3	0.3
Bottom friction ϵ	3×10^{-2}	0
Mixing ν	0	3×10^{-2}
Baroclinic damping ϵ_h	1×10^{-3}	1×10^{-3}
Resolution Δx	0.1	0.1
Time step Δt	0.2	0.2
Adams-bashforth ϵ_{AB}	0.1	0.1
Stommel layer $\frac{\epsilon}{\beta}$	0.1	0.1
Wave resolution $\frac{2S}{\Delta x}$	$\frac{1}{2}$	$\frac{1}{2}$

Table 5.1: Parameters for the double gyre integrations of the B grid, C grid and C_d models.

The behaviour of Rossby waves in the coupled C_d scheme can be obtained by following the procedure outlined at the beginning of this chapter. The zeroth order balance gives:

$$\hat{\mathbf{k}} \wedge \mathbf{u}_D^{(0)} = -S \nabla h \quad (5.32a)$$

$$\hat{\mathbf{k}} \wedge \mathbf{u}_C^{(0)} = -S \nabla \overline{h}^{xy} \quad (5.32b)$$

$$D_D^{(0)} = D_C^{(0)} = 0 \quad (5.32c)$$

At the next order, the vorticity equation is:

$$\frac{\partial}{\partial t} \zeta_D^{(0)} + D_C^{(1)} + \beta \overline{v^{(0)}}_C^y = -\hat{F}(\zeta_D)^{(0)} \quad (5.33a)$$

$$\frac{\partial}{\partial t} h + D_C^{(1)} = -\hat{H}(h) \quad (5.33b)$$

For the special case of $\hat{H} = \hat{F}$, the resulting dispersion relation is:

$$(\iota \hat{F} + \omega) = \frac{-\epsilon \beta S s_k c_k c_l^2}{1 + S(s_k^2 + s_l^2)} \quad (5.34)$$

which is that of the D grid. The D grid Rossby wave dispersion properties are like the C grid at high resolution, and a smoother version of the same at low resolution.

5.6 Shallow water model results

The B grid, C grid and C_d shallow water models were each integrated on a β -plane in a double gyre configuration. The parameters used are listed in table 5.1. The wave resolution was $\frac{2S}{\Delta x} = \frac{1}{2}$ (low resolution) so the B grid solution can be considered robust.

There are two aspects of the C_d method that need to be validated. The transient response or behaviour of waves has been analysed and can be verified by comparing the models in the process of adjustment or spin-up. The steady state solution also depends on the particular finite differencing and so comparison of the steady states should also be made.

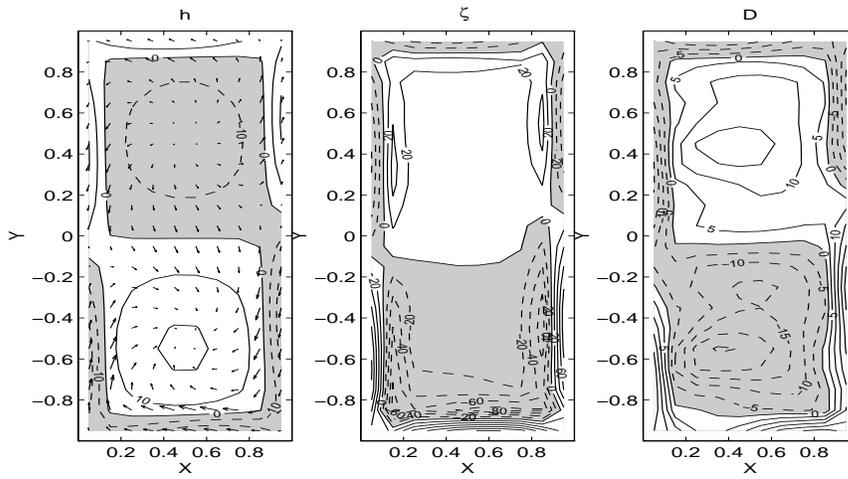
Figure 5.6 shows the solutions at $t = 40$, which is early in the spin-up of the integrations in experiment I. The C grid divergence and vorticity fields display the characteristic noisy response, and even the pressure field has some features that are not in the B grid solution. All the C_d fields are very similar to the B grid solutions in both structure and magnitude.

Figure 5.7 shows the steady state solutions for each model in experiment I. The solutions are all broadly similar. The divergence and pressure fields all have the same structure and magnitude. The boundary structure of vorticity is the only main difference. The C_d grid and B grid solutions are more similar with each other than with the C grid.

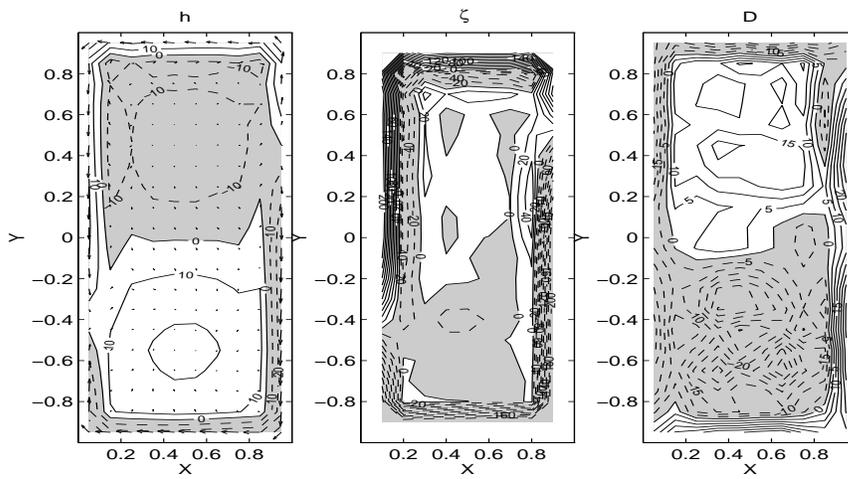
Experiment II was conducted to test that the success of the scheme was independent of the form of the dissipation in the momentum equations. Lateral mixing replaced the Newtonian damping term. Figures 5.8 and 5.9 show a

snapshot and the steady state as before. The noise level of the C grid is less apparent because of the scale selective dissipation. Nevertheless, it is still discernible that the C_d grid is in much better agreement with the B grid than is the C grid.

(B)



(C)



(C_d)

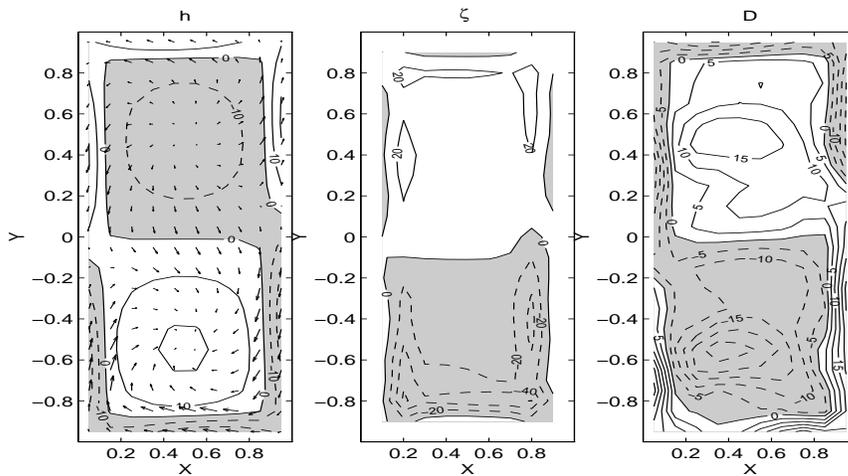
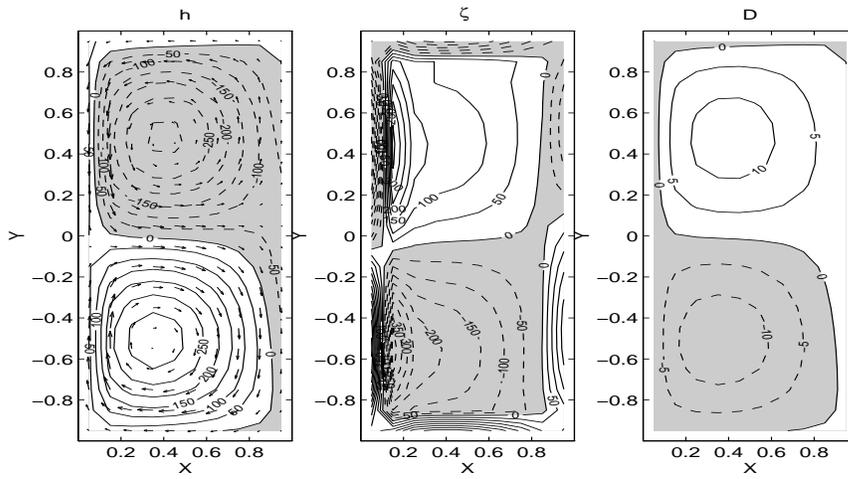
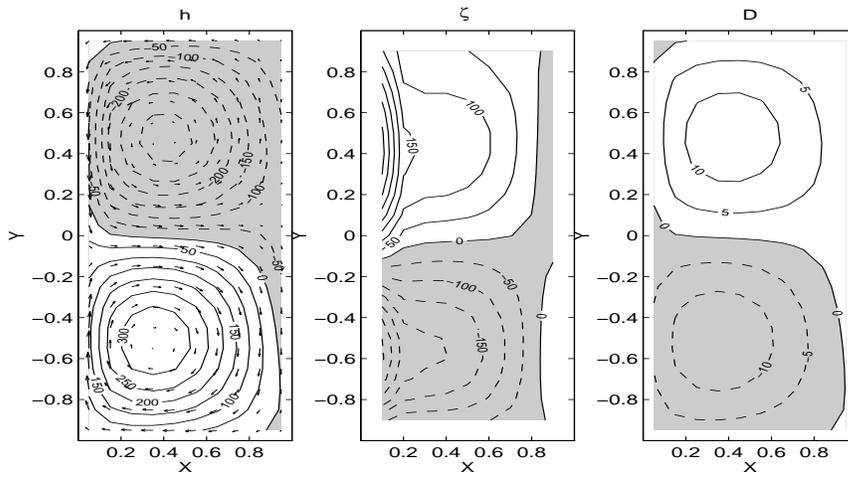


Figure 5.6: Surface elevation h with velocity vectors, vorticity ζ and divergence D for the three models B, C and C_d at $t = 40$ in experiment I. The solutions are in the early stage of adjustment. Note the noise in the divergence field of the C grid.

(B)



(C)



(C_d)

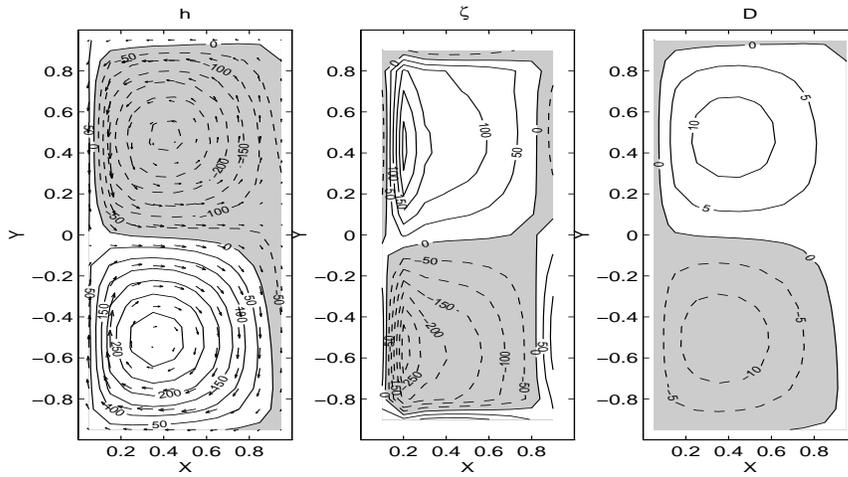
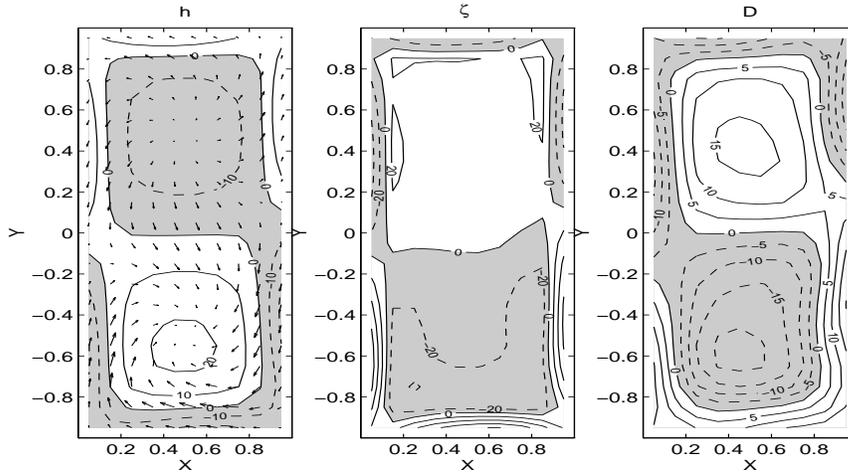
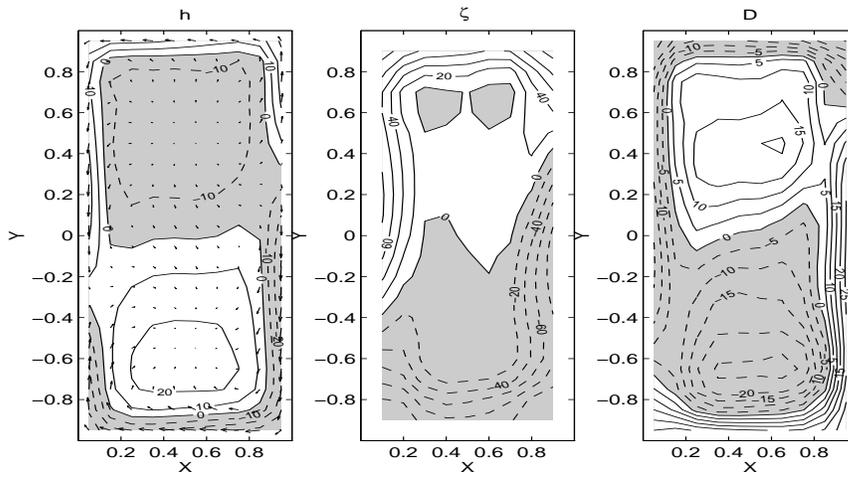


Figure 5.7: Surface elevation h with velocity vectors, vorticity ζ and divergence D for the three models B, C and C_d at $t = 5000$ in experiment I. The solutions are the steady state. The C grid vorticity field has a different boundary structure to the B and C_d grid solutions.

(B)



(C)



(C_d)

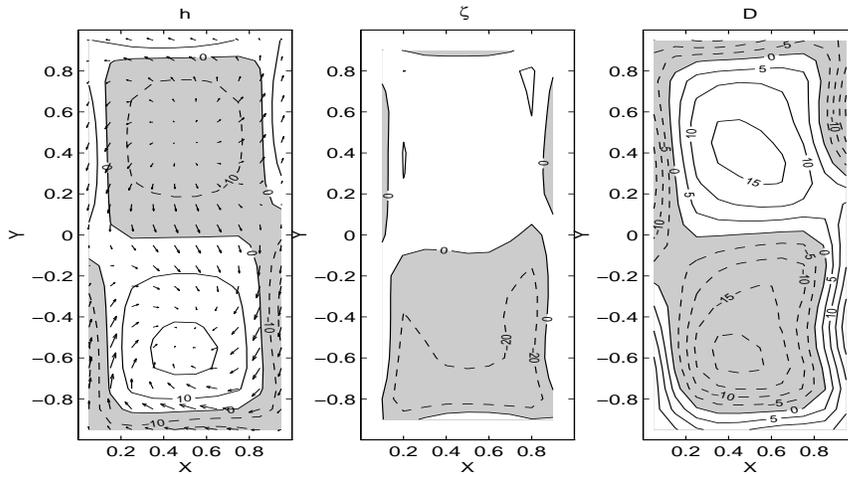
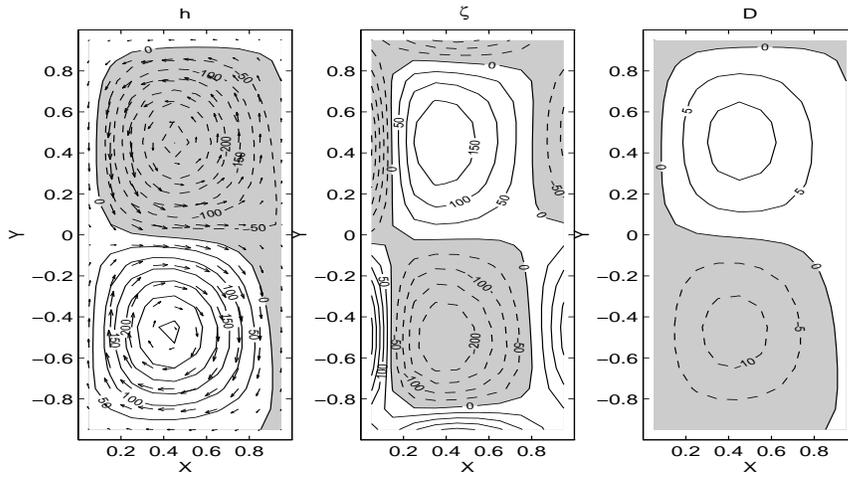
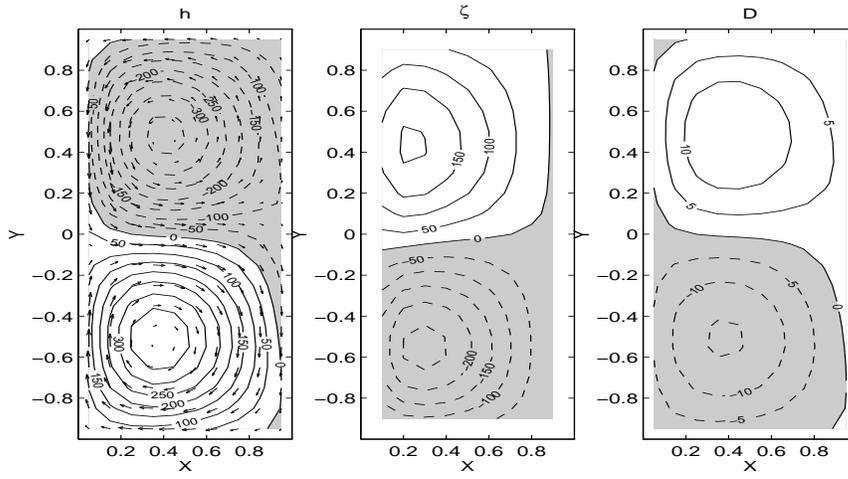


Figure 5.8: Surface elevation h with velocity vectors, vorticity ζ and divergence D for the three models B, C and C_d at $t = 40$ in experiment II. The solutions are in the early stage of adjustment. Note the substantial difference in pattern and magnitude of the vorticity field on the C grid.

(B)



(C)



(C_d)

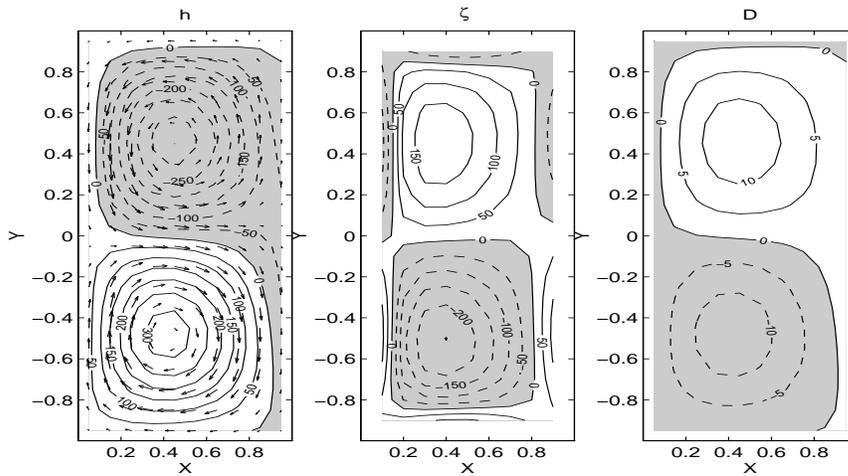


Figure 5.9: Surface elevation h with velocity vectors, vorticity ζ and divergence D for the three models B, C and C_d at $t = 5000$ in experiment II. The solutions have reached steady state.

5.7 Implementation of the C_d scheme in the GCM

The complete procedure applied to a baroclinic model will now be described. All terms excluding the pressure gradient and Coriolis terms are collected into a vector \mathbf{G} .

Define \mathbf{u}_C to fall at the integer time levels ($n-1, n, n+1$) and define the pressure variable to fall at the half-integer levels ($n-\frac{1}{2}, n+\frac{1}{2}$) as before. Now introduce new horizontal velocities \mathbf{u}_D to fall at the same half integer time levels as the pressure but are positioned in the D grid configuration.

At the beginning of the time-step to step forward for the $n+1$ C grid velocities, all variables at time $n-1\frac{1}{2}, n-1, n-\frac{1}{2}$ and n are known.

\mathbf{u}_D at $n+\frac{1}{2}$ is estimated by writing a centered single level time-stepping scheme in terms of \mathbf{u}_C and \mathbf{G} at time level n . Note that the pressure gradient terms have to be extrapolated forward in time in an analogous manner to the quasi-second order method AB2 (to avoid an implicit evaluation of the pressure at this stage). The zonal component of the D grid momentum equations is:

$$\begin{aligned} \frac{u_D^{(n+\frac{1}{2})}}{\Delta t} = \frac{u_D^{(n-\frac{1}{2})}}{\Delta t} &+ \overline{G_u^{(n)}}^{xy} + \frac{1}{\Delta x} \delta_x \left(\frac{3}{2} + \epsilon_{AB} \right) P^{(n-\frac{1}{2})} - \left(\frac{1}{2} + \epsilon_{AB} \right) P^{(n-1\frac{1}{2})} \\ &+ \overline{f^x} \left((1 + \epsilon_{AB}) v_C^{(n)} - \epsilon_{AB} v_C^{(n-1)} \right) \\ &+ \hat{F} \left(u_D - \frac{1}{2} (u_C^{(n)} + u_C^{(n-1)})^{xy} \right) \end{aligned} \quad (5.35)$$

The operator \hat{F} is the same as the dissipation operator used within the vector \mathbf{G} .

The C grid velocities are then found as in the conventional scheme except that now the Coriolis term is written as the vector \mathbf{C} :

$$\frac{u_C^{(n+1)}}{\Delta t} = \frac{u_C^{(n)}}{\Delta t} + \left(\frac{3}{2} + \epsilon_{AB} \right) G_u^{(n)} - \left(\frac{1}{2} + \epsilon_{AB} \right) G_u^{(n-1)} + \frac{1}{\Delta x} \delta_x P^{(n+\frac{1}{2})} + C_u \quad (5.36)$$

where C_u is defined:

$$C_u = \begin{cases} \overline{f \left(\frac{3}{2} + \epsilon_{AB} \right) v_C^{(n)} - \left(\frac{1}{2} + \epsilon_{AB} \right) v_C^{(n-1)}}^{yx} & \text{if C grid is used} \\ \overline{f^x} v_D^{n+\frac{1}{2}} & \text{if } C_d \text{ grid is used} \end{cases} \quad (5.37)$$

It should be apparent from this formulation that the D grid velocities only enter the model through the Coriolis term and are not used anywhere else. Obviously, if the C grid evaluation of Coriolis is used then the D grid velocities are redundant and need not be stepped forward.

The North Pacific integration in chapter 3 was used as the context in which to discuss problems inherent in the ‘C’ grid at coarse resolution. Figures 5.10 and 5.11 show the vertical velocity in the GCM for the same experiment but with the C_d scheme implemented. They correspond to figures 3.9 and 3.10 that were dominated by grid length waves in the unmodified C grid version of the model, just one month after the C_d scheme was switched off.

The scheme no longer allows grid scale noise to persist in the model. The major features of the Pacific circulation, such as the Kurishio and strong Equatorial upwelling, are unaffected by the scheme.

5.8 Discussion

Form of the Coupling During the development of the C_d scheme, it was not immediately realised that there was an optimum form for the coupling operator. Instead, the simplest possible form of Newtonian relaxation ($-\lambda(\mathbf{u}_D - \overline{\mathbf{u}_C}^{xy})$) was chosen. The relaxation time-scale was at first chosen using a somewhat subjective selection criteria. A connection between the grid-scale dissipation rates and the optimum choice of λ was empirically established. The implications of using such a coupling were that:

- all modes, both horizontal and vertical, would ‘feel’ the same damping rate. This could produce smoother than expected flows, with slower currents.
- for a poorly chosen λ , some modes could be excessively dissipated or tend to become de-coupled. The former manifests itself by smoothing the flow whilst the latter appears as a checker-board mode in the pressure field.
- there was an extra parameter for the model that had to be determined prior to integration.

All of these factors rendered this simple coupling method undesirable. The proposed method, of coupling using the dissipation operator, suffers from none of these problems.

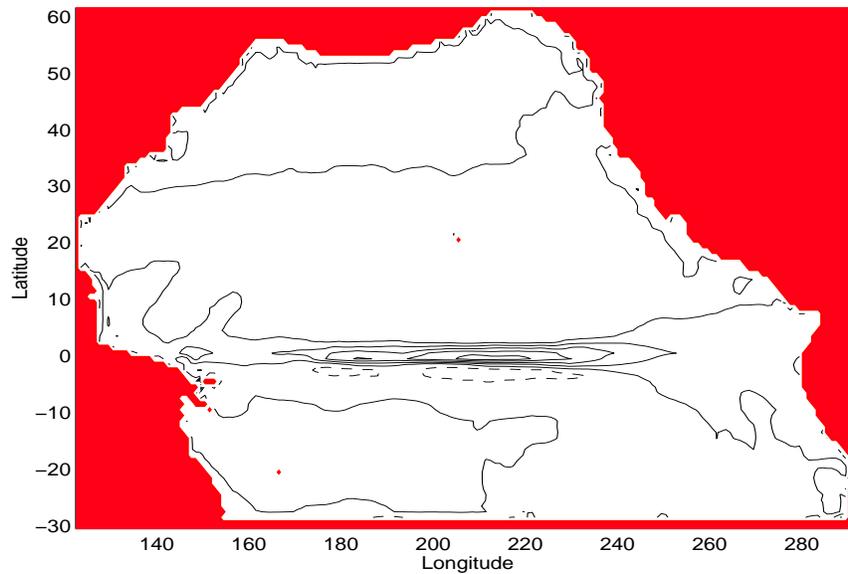


Figure 5.10: Vertical velocity at the base of the top layer in the model after one month. Contour interval is 0.01 mm s⁻¹.

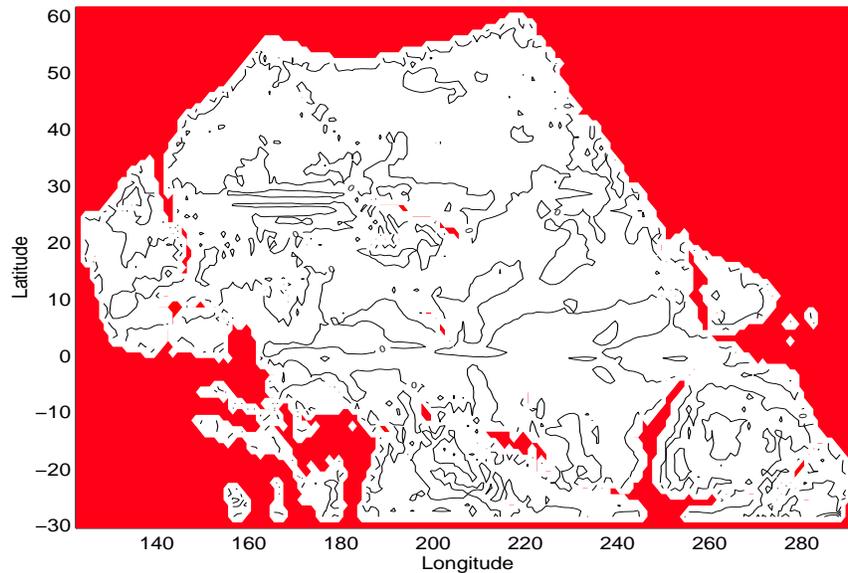


Figure 5.11: Vertical velocity at $z=-3200\text{m}$ in the model after one month. Contour interval is 0.05 mm s⁻¹.

Barotropic mode If a Newtonian relaxation term were used to do the coupling, then all modes (including the barotropic mode) would feel the damping effect. The proposed optimum coupling method (using the same form and strength as the dissipation in the momentum equations) is scale selective and has no additional damping effect on the circulation. A potential problem arises if the dissipation is predominantly in the vertical and the horizontal mixing is weak. The barotropic mode is only dissipated by horizontal mixing however and the slower rate means that the coupling is weaker. At 1 degree resolution, it is therefore quite possible to have a coupling rate of the order of months, using typical horizontal mixing parameters. The barotropic mode is bound to de-couple over this long period. The linear relaxation method would remove this de-coupling.

This eventuality has not been explored here but if it should be a problem, it is quite legitimate to separate the flow into barotropic and baroclinic components for the purpose of evaluating the Coriolis term. This could be done by making the separation just before evaluating Coriolis, and then re-combining the two components after the evaluation. The barotropic Coriolis term could be evaluated as on a conventional C grid since the barotropic deformation radius ($\sqrt{gH/f^2}$) is almost always resolved in ocean models.

Damping time-scales The avoid de-coupling of the two grids, the explicit dissipation in the momentum equations must be strong enough. For integrations on a β -plane (or sphere), the minimum dissipation required for stability is determined by the resolution. The Munk layer width goes like $\delta_{munk} \sim (\frac{A_h}{\beta})^{1/3}$ so for a given resolution, the lateral mixing must be larger than $A_h > \beta \Delta x^3$. The dissipation rate on the grid-scale is then $T_{\Delta x} \sim \frac{\Delta x^2}{A_h}$ which therefore goes like $T_{\Delta x} \sim (\beta \Delta x)^{-1}$. This gives a time-scale around six days at 1° resolution which is longer than the inertial period but shorter than any likely rate of de-coupling. At higher resolution, this imposed time-scale gets longer, but since, at higher resolution, the unmodified C grid is to be preferred and the C_d scheme can be switched off.

3-D Divergence on the D grid The shallow water equations are not exactly analogous to the primitive equations on which the GCM is based. The C grid velocities are non-divergent at each step. This is not true of the D grid velocities. If the D grid velocities become erroneously divergent it is conceivable that the stretching of vortex tubes may be in error. However, in practice, the D grid velocities have been close enough to the C grid that the difference is insignificant. The vorticity balance was modelled well enough in the experiments described.

Energetics The energetics of the scheme are far from satisfactory. Inspection of the Coriolis term reveals that the scheme can *not* conserve energy. The Coriolis acceleration should be perpendicular to the velocity. The velocities used to evaluate that acceleration are held on the D grid. The kinetic energy equation involves the scalar product of the D grid velocities with the C grid velocities, and should vanish according to the continuous analogue. The two velocities on C and D grids might be similar, but it is not guaranteed that they are the same and so the Coriolis acceleration and C grid velocity are not necessarily perpendicular. *i.e.* $\mathbf{u}_D \wedge \mathbf{u}_C \neq 0$.

Despite this apparent inability to conserve energy, actual integration of the scheme shows that energy leakage is quite minimal. The only reason for this is that on time-scales longer than f^{-1} , the D grid velocities are very close to the C grid velocities, as intended. Extended integrations of the shallow water models always reached a steady state and extended integrations of the GCM have not revealed any energetic problems.

Time-step limitation The method is fully explicit in Coriolis and therefore has a time-step limitation of 1.5 hours. This limitation is set by the inertial frequency and the only way to avoid this is to move to an implicit formulation. Coriolis on a C grid with conventional implicit time-stepping involves the substantial modification of the elliptic operator that appears in the pressure equation. Here, this is most likely also true if the C_d scheme were evaluated in some implicit manner. It is possible that the D grid velocities could be integrated in a split time-stepping manner, using a small time-step for the D grid, and a long time step for the C grid. This may have some computational advantages sine the D grid equations are linear and fully explicit.

5.9 Summary

An analysis of Rossby wave behaviour in numerical models showed that the gridding determines the resulting dispersion properties. No one grid has ideal behaviour. The best grid, in terms of inertia-gravity wave and Rossby wave propagation, is the Eliassen DD' grid.

A numerical scheme (C_d), based on the time-staggering method used in the Eliassen DD' grid, is developed. The resulting scheme results in well behaved inertia-gravity and Rossby waves on a C grid. The method was designed so that it should involve only minor changes to an existing C grid model. The computational cost of the scheme is minimal.

Comparison of the C_d scheme with the B and normal C grids is conducted with numerical shallow water models. The results show that at low-resolution, where the C grid fails, the C_d scheme behaves as well as the B grid.

The C_d scheme introduces two new variables; a second pair of horizontal velocity components carried on a D grid. They are integrated forward in parallel with the rest of the model and affect the model only through the evaluation of the Coriolis term. The D grid velocities are coupled to the C grid by means of a scale selective relaxation. The justification for the whole procedure is that it selectively retains the beneficial aspects of the C grid whilst improving the representation of the Coriolis term.

The C_d scheme was implemented in the GCM. Extended integrations for the North Pacific integration show that the scheme has clearly remedied the inertia-gravity wave problem associated with the spatial averaging of the Coriolis term on the C grid.

Chapter 6

Representing Topography: the finite volume approach

Topography in atmospheric models has often been treated by use of σ co-ordinates (sigma) which follow the terrain. The method is well suited to smoothly varying, low-lying topography. It can become hydrostatically inconsistent if the slope is of the order or larger than the aspect ratio of model grid spacing. Special consideration of the horizontal discretisation must be made to avoid this. A major failing of the σ co-ordinate system is that the inaccuracies arising from the projection of gravity onto the horizontal components of the model spontaneously generate motion. That is, a resting ocean, in the presence of sloping bottom, may begin to move. The motion is driven by truncation errors in the evaluation of horizontal pressure gradients on sloping co-ordinate surfaces.

Terrain following co-ordinates do ease the representation of boundary conditions but at the same time they tend to complicate the form of the elliptic problem; cross-terms appear due to the projections onto the transformed system of the horizontal and vertical. Nevertheless, terrain following co-ordinates have successfully been used in ocean modelling (see Haidvogel et al. [HWY91] and Mellor [Mel92]), though the application of such models has been limited to regional studies. In the context of a general purpose circulation model, physical co-ordinates tend to be a more direct and robust approach.

The earliest attempts to model the world ocean circulation either assumed a flat bottomed ocean (see Takano, 1974 [Tak74]) or approximated the topography by choosing layer depths to coincide with the bottom depth. (see Bryan, 1969 [Bry69]).

Figure 6.1 illustrates some of the main methods of representing topography. The simplest, though coarsest, representation is the step-topography (top-left) used in the Bryan code. The number of layers used in the vertical is a function of horizontal position, and the depth of the ocean is truncated to “fit” the model layers. A simple algorithm to determine whether a cell is land or water is to measure the fraction of land/water in the real bathymetry and either empty or fill it in according to whether it is more than or less than $\frac{1}{2}$ land. This conserves the volume of ocean only in an approximate statistical sense.

An obvious improvement (top-right) on standard step-wise representation is to allow the thickness of the bottom layer to vary (see Semtner and Mintz, 1977 [SM77]). The representation is still step-wise, though the volume of the water column is better represented and thus the volume of ocean is too. However, such an approach was deemed an unwanted complication in the model formulation since the variable vertical resolution of the bottom layer required extra storage and careful consideration of the finite differencing. These objections no longer apply due to recent advances in computer technology.

Here, a finite volume approach will allow variations in the lateral and vertical extent of cells neighbouring a solid boundary. In principle, the volume and associated facial areas that define a 3-D grid of boxes, overlaid on high-resolution topography, can be used. In practice, a minimum volume size must be imposed to maintain stability. This is done by rounding the volume of the cell to some small number. The stability requires simply that a volume not be too small that the effective grid-size violate a CFL criterion.

In the following sections, the finite volume approach will be described and applied to the incompressible Navier-Stokes equations. The resulting model is then tested by modelling two distinct problems of oceanic relevance and interest; barotropic flow over a tall sea-mount and simulation of the topographic β effect in a homogeneous layer of fluid.

The results from these experiments suggest that the finite volume method has some advantages over the conventional methods mentioned above. Its use, however, may demand a smaller time-step to satisfy a more stringent CFL criteria.

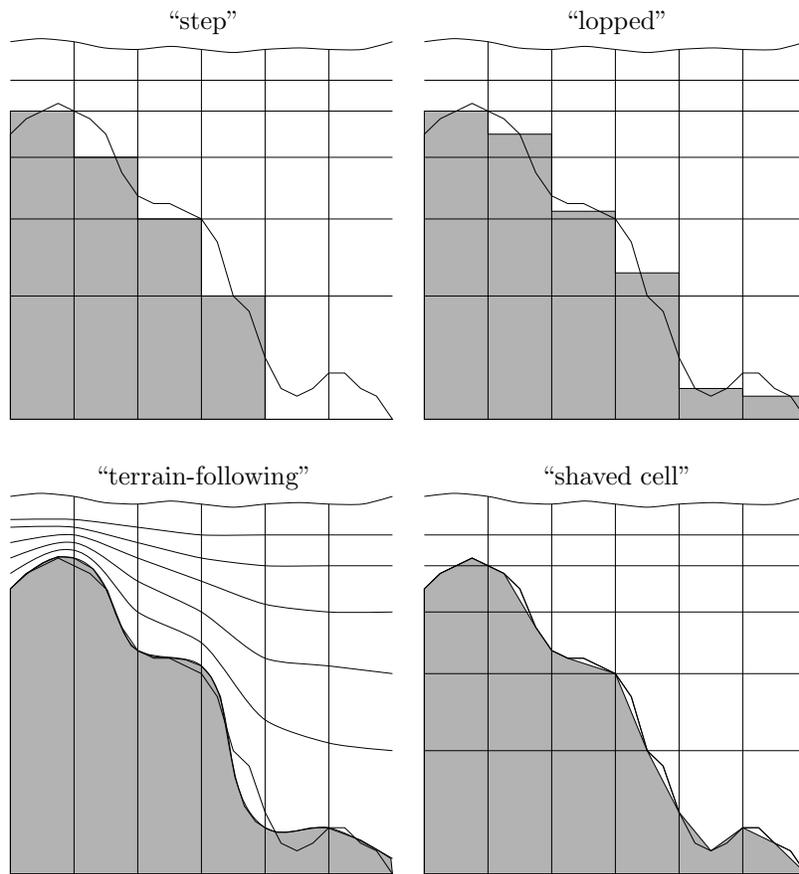


Figure 6.1: Four methods of representing topography in numerical models. The shaded region represents the land mask as seen by each model. The cross-gridded lines are the model cells defined by the co-ordinate system. Top-left: step-wise representation of topography where the ocean depth is truncated to fit the model layer depths. Top-right: A variation on step-wise topography “lops” the bottom layer to better approximate the volume of a column of ocean. Bottom-left: σ co-ordinates use a terrain following co-ordinate transformation such that the model domain is regular. Bottom-right: The proposed finite-volume approach can, in principle, take account of the observed volume of each column in the ocean.

6.1 Formulations in numerical modeling

A numerical model attempts to represent a continuous system by some finite set of variables and algebraic equations. The state of the system and the governing equations are truncated. The finite difference and finite element techniques approach the formulation process from different ends of the spectrum.

In finite differencing, the primary objective is to reduce the truncation error of the discrete equations. A brief overview of finite difference methods was given in section 3.1 and a description of short-hand notation was out-lined in section 3.2. In the Galerkin approach (of which finite element methods are one), the focus is on the representation of the state of the system. The governing discrete equations are then deduced from this truncated representation of the state.

In many cases, the resulting algebraic equations derived using finite difference and finite element techniques have many similarities and, indeed, can be identical. The interpretation of the discrete model variables (ie. representation of the state) in the two methods can be very different, however.

The finite volume technique, as described in this chapter, can be considered to lie between these two methods. It is equally concerned both with the truncation error of the governing equations and with the representation of the state. As in finite element and finite difference representations, the finite volume approach can yield a very similar final algorithm if applied to a domain which is uniformly gridded.

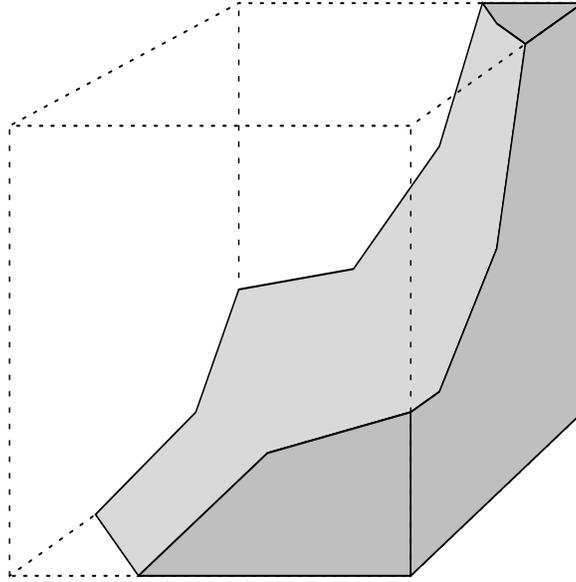


Figure 6.2: A schematic of a shaved cell. The regular (unshaved) cell defined by the grid is depicted by the dotted lines. The solid surface of the topography intersects the volume changing the shape/size of the volume of fluid and the six areas through which fluxes are exchanged with neighbouring cells. Despite the apparent complexity in shape of the shaved cell, there are still only seven attributes of the cell that are needed to evaluate the net accumulation of fluxes into the volume; the volume and six areas of the *open* cell/cell interfaces.

6.2 Finite volume method

The concept behind the finite volume technique is to integrate the principle governing equations over finite volumes into which the model has been divided. The resulting integrated expressions involve either volume averages of the state variables or the area integrals of flux incident at the surface of the volume (by the Gauss divergence theorem). It is these expressions of volume/area average quantities that are used to define the representation of the state of the system.

The ocean is first carved up into consecutive finite volumes of the form of hexahedrons (boxes). The volumes fit tightly together and have their faces aligned such that the normal vectors are pointed along one of the independent orthogonal co-ordinate axis. The volumes are quasi-regular in that looking along one co-ordinate, the separation of faces normal to that co-ordinate is invariant along the remaining two axis. ie. zonal grid-spacing can vary in x only, meridional grid-spacing can vary in y only and vertical spacing varies in z only. This is the only arrangement of variable grid-spacing that allows each face of a volume to be aligned with the orthogonal co-ordinate system.

The nodes that are nominally at the centres of the volumes are labeled i, j, k and the volume itself designated $V_{i,j,k}$. Each face defining a hexahedron is shared between two consecutive volumes. The eastern face of volume $V_{i,j,k}$ has area $A_{i+\frac{1}{2},j,k}^u$, the northern face has area $A_{i,j+\frac{1}{2},k}^v$ and the upper face has area $A_{i,j,k+\frac{1}{2}}^w$.

Solid boundaries that intersect an otherwise regular cell modify the volume of fluid and the areas of the *open* cell/cell interfaces as depicted in figure 6.2. The surface area of the solid interface never enters the problem. The alignment of the *open* faces with the orthogonal coordinate system is not affected by the introduction of topography. Mean quantities within the volume evolve according to the net accumulation of fluxes through the open faces. The fluxes in the model will be defined to be the mean value over the open area. If a face is entirely closed off, this is simply represented as the area of the face vanishing. The boundary condition of no flux through solid boundaries is implicitly met by budgeting only with fluxes through open areas.

The pivotal equation in the incompressible Navier-Stokes equations, concerning the finite volume approach, is the continuity equation (incompressibility condition). The continuity equation expresses non-divergence of the flow; zero net accumulation of volume flux. In other words, integrating the continuity equation over a finite volume expresses the fact that the volume itself is constant.

In the following sections, the finite volume approach will be applied to the model equations as described in chapter 2. The method will be implemented by adapting a finite-difference code. One aspect of this finite-difference code is that the algorithm used to solve the elliptic problem is the pre-conditioned conjugate gradient method. As

implemented, the method is limited to solving positive definite, symmetric matrix problems. This requirement will limit the form that the finite volume momentum equations can take because the finite volume approach can otherwise lead to an asymmetric matrix problem. We therefore apply the finite volume approach to the continuity and tracer equations but use a less general approach when forming the finite volume momentum equations. The form of the momentum equations will be found that both leads to a symmetric elliptic problem and conserves energy.

6.2.1 Continuity equation and boundary conditions

The continuity equation is the pivotal equation when applying the finite volume approach to the incompressible Navier-Stokes equations. Integrating continuity over the volume $V_{i,j,k}$ yields:

$$\iiint_{V_{i,j,k}} \nabla \cdot \mathbf{u} \, dV = \iint \mathbf{u} \cdot \hat{\mathbf{n}} \, dA = 0 \quad (6.1)$$

No normal flow is imposed at solid interfaces, and at the rigid-lid at the surface. This is naturally applied in the integral divergence form of the continuity equation as written above. The elliptic equation for pressure, that arises from substituting the momentum equations in to the incompressibility condition as written, automatically has the appropriate boundary conditions if use is made of 6.1.

The integrated expression, $\iint \mathbf{u} \cdot \hat{\mathbf{n}} \, dA$, involves only fields evaluated on the surface of the volume. Divergence within the interior of the volume is budgeted for implicitly. The area integrals applied to the normal component of velocity can be written in terms of the area and mean value of the velocity component. For example, the eastern face contribution to the budget is:

$$\iint_{A_{i+\frac{1}{2},j,k}^u} u \, dy \, dz \equiv A_{i+\frac{1}{2},j,k}^u u_{i+\frac{1}{2},j,k} \quad (6.2)$$

which is simply the definition of the mean value of u on the eastern face multiplied by its area $A_{i+\frac{1}{2},j,k}^u$.

The integral form of the continuity, for the finite volumes defined above, then takes the form:

$$\begin{aligned} & A_{i+\frac{1}{2},j,k}^u u_{i+\frac{1}{2},j,k} - A_{i-\frac{1}{2},j,k}^u u_{i-\frac{1}{2},j,k} \\ + & A_{i,j+\frac{1}{2},k}^v v_{i,j+\frac{1}{2},k} - A_{i,j-\frac{1}{2},k}^v v_{i,j-\frac{1}{2},k} \\ + & A_{i,j,k+\frac{1}{2}}^w w_{i,j,k+\frac{1}{2}} - A_{i,j,k-\frac{1}{2}}^w w_{i,j,k-\frac{1}{2}} = 0 \end{aligned} \quad (6.3)$$

where the averaging operator has been dropped from the velocities. In short-hand difference notation this is:

$$\delta_x(A^u u) + \delta_x(A^v v) + \delta_z(A^w w) = 0 \quad (6.4)$$

In order to satisfy continuity exactly, the momentum variables in the model are appropriately defined as averages over respective faces of the volume.

If this is so, then it should be understood that the model variable $u_{i+\frac{1}{2},j,k}$, for example, contains no information about the continuous u field within the volume itself. In the x direction, the representation of the continuum u field looks like a series of delta functions at discrete intervals. Much as in finite differencing, the y - z mean of continuum u matches the model variables at these discrete nodes in the x direction.

Although the foundation of a discretised model have been laid, it should be emphasized that at this stage, in satisfying the integral form of continuity, the discrete model variables are an exact representation of the continuum. There is no truncation error in the evaluation of continuity so long as the model velocities are defined to be the means over their respective areas.

6.2.2 Tracers

Just as continuity is succinctly handled in the finite volume approach, conservation of tracers is likewise naturally represented. In an incompressible fluid, advection of tracer concentrations (salinity, temperature, etc...) can be written as the divergence of advective fluxes (cf. salt conservation):

$$\frac{\partial}{\partial t} S + \nabla \cdot (S \mathbf{u}) = Q_s \quad (6.5)$$

where Q_s represents real sources and sinks (such as molecular diffusion, riverine discharge, chemical reactions, etc...) of the passive tracer S . Integrating over the same finite volume, $V_{i,j,k}$, as was used for continuity, gives:

$$\frac{\partial}{\partial t} \iiint_{V_{i,j,k}} S dV + \iint \mathbf{u} S \cdot \hat{\mathbf{n}} dA = \iiint_{V_{i,j,k}} Q_s dV \quad (6.6)$$

The first term suggests that the volume mean of the tracer is a suitable discrete variable. This has the facility of unambiguously and directly budgeting for the total amount of tracer in the model, though conservation of higher moments might need further consideration.

The divergence of advective fluxes can be evaluated much as described previously for the continuity equation. Here, however, the flux is the area average of a product of other model variables as opposed to the product of area averaged model variables ($\overline{vS} = \bar{v}\bar{S} + \overline{v'S'}$). This integral must be expanded and expressed in terms of available quantities. This will require the parameterization of an *eddy* flux, as will be shown later.

Evaluation of the source term is dependent upon the particular tracer being considered and will not be dealt with in full detail. In general, most sources and sinks can be written as the divergence of a flux, the flux often being a boundary condition on the model associated with the forcing.

Returning to the advection term, the area integral of the flux through the eastern face can be written:

$$\iint_{A_{i+\frac{1}{2},j,k}^u} uS dy dz = A_{i+\frac{1}{2},j,k}^u u_{i+\frac{1}{2},j,k} \bar{S}^{A^u} + \iint_{A_{i+\frac{1}{2},j,k}^u} u'S' dy dz \quad (6.7)$$

where the primes denote deviations of the continuum field from the mean value on the face.

There are two issues of closure here; one involves expressing \bar{S}^{A^u} in terms of the volume averaged variable $S_{i,j,k}$ and the other requires parameterization of the *eddy* flux $\overline{u'S'^{yz}}$. The latter has historically been an important issue and many parameterization schemes have been developed to express $\overline{u'S'^{yz}}$ in terms of the mean quantities. These should be applicable here even though traditionally the eddy flux is averaged over three spatial dimensions and time.

A finite difference approach of minimizing the truncation error in evaluating \bar{S}^{A^u} could be used which would favour some high order interpolation scheme. However, a property of the continuous system is that it conserves higher moments under advection. Here, conservation properties are deemed more important than accuracy since the principle behind the finite volume approach is one of conservation of volume.

Suppose the advection terms are differenced as:

$$V \frac{\partial S}{\partial t} + \delta_x (A^u u \bar{S}^x) + \delta_y (A^v v \bar{S}^y) + \delta_z (A^w w \bar{S}^z) \quad (6.8)$$

then conservation of the second moment can be shown as follows.

$$\begin{aligned} & S \left(\delta_x A^u u \bar{S}^x + \delta_y A^v v \bar{S}^y + \delta_z A^w w \bar{S}^z \right) \\ &= S^2 (\delta_x A^u u + \delta_y A^v v + \delta_z A^w w) \\ & \quad + S \left(\overline{A^u u \delta_x S^x} + \overline{A^v v \delta_y S^y} + \overline{A^w w \delta_z S^z} \right) \\ &= S^2 (\delta_x A^u u + \delta_y A^v v + \delta_z A^w w) \\ & \quad + \overline{\bar{S}^x A^u u \delta_x S^x} + \overline{\bar{S}^y A^v v \delta_y S^y} + \overline{\bar{S}^z A^w w \delta_z S^z} \\ & \quad - \frac{1}{4} (\delta_x (A^u u (\delta_x S)^2) + \delta_y (A^v v (\delta_y S)^2) + \delta_z (A^w w (\delta_z S)^2)) \\ &= \frac{S^2}{2} (\delta_x A^u u + \delta_y A^v v + \delta_z A^w w) \\ & \quad + \delta_x \left(A^u u \frac{1}{2} \wr S_x^2 \right) + \delta_y \left(A^v v \frac{1}{2} \wr S_y^2 \right) + \delta_z \left(A^w w \frac{1}{2} \wr S_z^2 \right) \end{aligned}$$

where $\wr S_x^2$ is the geometric product between neighbouring nodes. The notation is described in section 3.2.

The first contribution vanishes since the advecting flow is non-divergent. The second is the discrete analogue of $\nabla \cdot \frac{1}{2} S^2 \mathbf{u}$ and has no net contribution when summed over the model.

In difference notation, the full form of the tracer equation is:

$$\frac{\partial S}{\partial t} + \frac{1}{V} \left(\delta_x A^u (u \bar{S}^x + \overline{u'S'}) + \delta_y A^v (v \bar{S}^y + \overline{v'S'}) + \delta_z A^w (w \bar{S}^z + \overline{w'S'}) \right) = Q \quad (6.9)$$

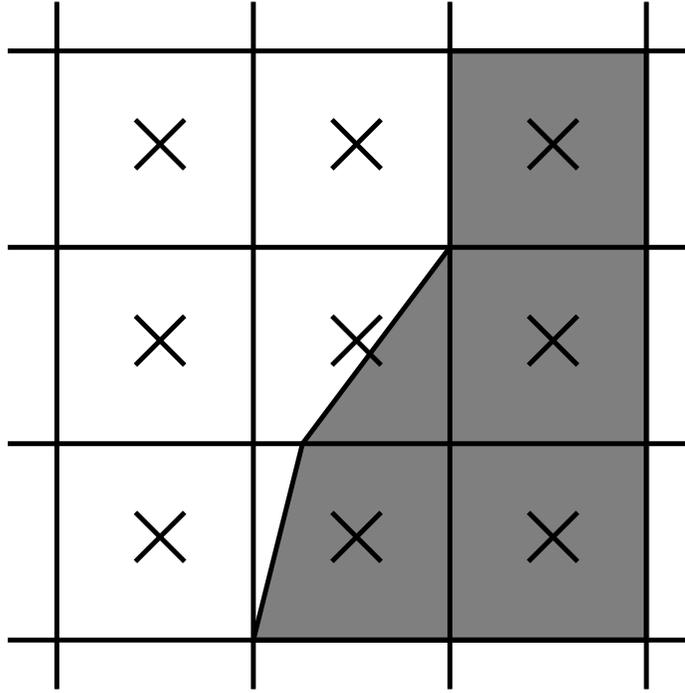


Figure 6.3: The shaving of cells at a boundary causes difficulties in defining the centre of a volume for finite difference purposes. In the top row, all the both water-filled cells are regular. The centre of a volume is well defined and the gradient of properties at the interface between cells easily defined. For the middle and bottom rows this definition appears to break down. The bottom row in particular shows the regular centre of a cell apparently within solid land.

The divergence of eddy fluxes must obviously conserve the total amount of tracer. Conservation of the second moment is conditional on the particular parameterization of sub-grid scale processes. A commonly used parameterization of the sub-grid scale eddy flux is as a down gradient flux. For example, the zonal eddy flux would be parameterised and evaluated as:

$$\overline{u'S'} \approx -\kappa \frac{\partial S}{\partial x} \approx -\frac{\kappa}{\Delta x} \delta_x S \quad (6.10)$$

Note that no weighting by volumes is used since the gradient should be a function of the difference in model variables only.

6.2.3 Momentum Equations

The momentum equations are integrated over finite volumes (labeled V^u , V^v and V^w) that are centered on the faces of the pressure/tracer volume ($V_{i,j,k}$) used earlier.

In this study, we had to limit the form that V^u , V^v and V^w could take for the following reason. The finite volume method will later be tested by adapting a finite difference code in which the elliptic problem is solved using a pre-conditioned conjugate gradient algorithm. This algorithm will only work if the problem is symmetric. To guarantee symmetry, discretisation of the pressure gradient will be considered first. An arbitrary choice of V^u , V^v and V^w will destroy the symmetry of the elliptic operator. The form of V^u , V^v and V^w is consequently chosen with care such that it retains symmetry of the elliptic problem and also conserves energy. Thereafter, the discretisation of Coriolis and advection terms will be found so that energy is conserved, bearing in mind the special form of the volumes V^u , V^v and V^w .

Pressure gradients

One might evaluate the pressure gradient by taking the difference between the mean pressures in two neighbouring volumes and dividing by the separation of the *centre* position of the volumes. This makes sense for regularly shaped volumes, but if, as is intended, the volumes acquire irregular shapes, then the centre becomes a little more difficult to define.

Figure 6.3 shows a schematic of some arbitrary topography or coast-line. The grid of volumes, or cells, is regular. The boundary cells have been shaved to represent the topography. For the top row, the horizontal gradient of

properties between water-filled cells is apparently easily defined by the separation of the centres of the cells. The centres are less obviously defined in the middle and bottom rows. The bottom row in particular has the regular cell centre within land.

In the continuous formulation, a diagnostic elliptic equation for the pressure was found from continuity and the momentum equations. It takes the form:

$$\nabla \cdot \mathbf{F} = \nabla \cdot \mathbf{G} - \frac{\partial}{\partial t} \nabla \cdot \mathbf{u} \quad (6.11)$$

where the last term vanishes due to the incompressibility condition. The left hand term becomes the elliptic operator (∇^2) on substituting $\mathbf{F} = \nabla p$ with the right hand side providing the source.

As written, the problem has an infinite set of solutions \mathbf{F} that satisfy the lateral boundary conditions since a non-divergent vector field can be added to \mathbf{F} . However, the elliptic problem, $\nabla^2 p = \nabla \cdot \mathbf{G}$, can be equivalently expressed:

$$\nabla \cdot \mathbf{F} = \nabla \cdot \mathbf{G} \quad (6.12a)$$

$$\nabla \wedge \mathbf{F} = \mathbf{0} \quad (6.12b)$$

This is a well posed problem that has a unique solution [GB79]. The very fact that system 6.12 is equivalent to the elliptic equation for pressure *and* has a unique solution suggests that the particular representation of \mathbf{F} in a discrete sense does not change the solution, so long as that discretisation satisfies the zero curl condition.

Since it is \mathbf{F} that ensures the non-divergence, rather than the particular discretisation of \mathbf{F} , there is some flexibility in its choice of discretisation. Therefore, for the convenience of creating a symmetric elliptic problem, a fixed grid spacing is used to evaluate the pressure gradient.

This might seem to ignore the role of the boundaries and boundary conditions in the elliptic problem. However, as was mentioned in the context of the continuity equation, the boundary conditions are applied to the continuity equation in the process of integrating it over finite volumes. Shaving cells modifies the areas and volumes of the cells which are explicitly required in the resulting elliptic problem.

The simplest, discrete form of the pressure gradient, that gives rise to a symmetric elliptic problem, is used:

$$\frac{\partial u}{\partial t} + \frac{1}{\rho_o} \frac{1}{\overline{\Delta x}} \delta_x p \quad (6.13a)$$

$$\frac{\partial v}{\partial t} + \frac{1}{\rho_o} \frac{1}{\overline{\Delta y}} \delta_y p \quad (6.13b)$$

$$\frac{\partial w}{\partial t} + \frac{1}{\rho_o} \frac{1}{\overline{\Delta z}} \delta_z p \quad (6.13c)$$

Energy conservation requires an analogue of the property:

$$\mathbf{u} \cdot \nabla p = \nabla \cdot (p\mathbf{u}) - p\nabla \cdot \mathbf{u} = \nabla \cdot (p\mathbf{u}) \quad (6.14)$$

which means that the pressure gradient should not contribute to the global energy budget.

The discrete form of 6.14 is:

$$\begin{aligned} \frac{\overline{V^u u}}{\Delta x} \delta_x p &+ \delta_x \left(\overline{p^x} \frac{V^u u}{\Delta x} \right) &+ p \delta_x \left(\frac{V^u u}{\Delta x} \right) \\ &+ \frac{\overline{V^v v}}{\Delta y} \delta_y p &= \delta_y \left(\overline{p^y} \frac{V^v v}{\Delta y} \right) &- p \delta_y \left(\frac{V^v v}{\Delta y} \right) \\ &+ \frac{\overline{V^w w}}{\Delta z} \delta_z p &+ \delta_z \left(\overline{p^z} \frac{V^w w}{\Delta z} \right) &+ p \delta_z \left(\frac{V^w w}{\Delta z} \right) \end{aligned} \quad (6.15)$$

where use of 3.12e has been made.

The last term should vanish by continuity and this can only happen if:

$$V^u = \overline{\Delta x}^x A^u \quad (6.16a)$$

$$V^v = \overline{\Delta y}^y A^v \quad (6.16b)$$

$$V^w = \overline{\Delta z}^z A^w \quad (6.16c)$$

Equations 6.16 exclude the possibility of using arbitrary volumes for the velocity variables, even where the volumes abut a solid wall. For the scheme to be energy conserving, the effective volumes for velocity variables must be chosen as in 6.16. However, the effects of topography are still reflected in the volume and face areas used in the continuity and tracer equations.

In review, the form of the pressure gradient terms is restricted by a computational preference for a symmetric elliptic problem. The finite volumes used in the discrete momentum equations must consequently take a particular form to conserve energy.

Non-linear terms

The form of the advection scheme is determined by considering the conservation properties of the advection operator in the continuum. Total kinetic energy should be conserved and this is only possible if the advecting fluxes are the non-divergent as defined by the continuity equation.

Suppose, the terms in the zonal momentum were differenced as:

$$\frac{\partial}{\partial t} u + \frac{1}{V^u} (\delta_x(U^* \bar{u}^x) + \delta_y(V^* \bar{u}^y) + \delta_z(W^* \bar{u}^z)) \quad (6.17)$$

where U^* , V^* and W^* are area integrated velocity components. To conserve $\sum V^u u^2$, the advecting fluxes, U^* , must be non-divergent according to $\delta_x U^* + \delta_y V^* + \delta_z W^* = 0$. The advecting fluxes must therefore be defined:

$$U^* = \overline{A^u u^x} \quad V^* = \overline{A^v v^x} \quad W^* = \overline{A^w w^x} \quad (6.18)$$

Notice that $A^u u$, $A^v v$ and $A^w w$ are recurring quantities in the continuity equation and tracer advection.

The discrete form for the advection terms in the zonal momentum equation is:

$$\frac{\partial}{\partial t} u + \frac{1}{V^u} (\delta_x(\overline{A^u u^x} \bar{u}^x) + \delta_x(\overline{A^v v^x} \bar{u}^y) + \delta_x(\overline{A^w w^x} \bar{u}^z)) \quad (6.19)$$

where it should be remembered that $V^u = A^u \overline{\Delta x^x}$.

Coriolis terms

Integrating the $\sin \phi$ Coriolis terms over the volumes of the respective velocities variables gives:

$$V_{i+\frac{1}{2},j,k}^u \frac{\partial}{\partial t} u_{i+\frac{1}{2},j,k} - \iiint f v \, dV \quad (6.20a)$$

$$V_{i,j+\frac{1}{2},k}^v \frac{\partial}{\partial t} v_{i,j+\frac{1}{2},k} + \iiint f u \, dV \quad (6.20b)$$

The variables u and v are carried at different grid-points which means that they must be spatially interpolated in the horizontal directions. f is normally defined at p latitudes and so must be included within the interpolation when being used at v points¹. In order to be energy conserving, some volume common to the neighbouring u - v points must be used. The most obvious volume to use in each interaction between four neighbouring u - v points is the p volume $V_{i,j,k}$.

An energy conserving form for the discrete Coriolis terms is:

$$V^u \frac{\partial}{\partial t} u - \overline{f V \bar{v}^y} \quad (6.21a)$$

$$V^v \frac{\partial}{\partial t} v + \overline{f V \bar{u}^x} \quad (6.21b)$$

As written, the Coriolis terms do not generate spurious volume integrated kinetic energy. For regular, unshaved cells, the scheme is that described by Arakawa and Lamb, 1977 [AL77, HW80]. Further, f can vary in any direction which is potentially useful for global integrations in which the poles might want to be shifted².

Use of the C_d scheme for evaluating the Coriolis terms in the finite volume model destroys the ability to identically conserve energy. Since energy conservation is a property on which the discretisation of the momentum equations is based we prefer to not employ the C_d scheme when using the finite volume code. The C_d scheme was therefore not used in any of the experiments even though a crude implementation of the scheme has been made by using straight, unweighted interpolation where required.

¹Taking f outside the interpolation operator and interpolating it separately, spuriously generates kinetic energy. See Arakawa and Lamb, 1977 [AL77].

²Shifting poles into solid land reduces the restrictive CFL criteria resulting from high zonal resolution near the poles

6.2.4 Numerical stability

The linear numerical stability of each individual term in the model can be analyzed. This does not guarantee overall stability but it is invariably a good guide for choosing the model and numerical parameters.

The quasi-second order Adams-Bashforth scheme (ABII) is used for all terms except the pressure gradient. The analysis of this scheme is complicated but it suffices to satisfy a CFL criteria similar to a forward in time discretisation for non oscillatory terms.

On a regular grid, a time-lagged diffusion term (forward in time, centered in space - FTCS) is differenced as:

$$\frac{1}{\Delta t}(\phi^{n+1} - \phi^n) = \frac{\kappa}{\Delta x^2}(\phi_{i-1}^n - 2\phi_i^n + \phi_{i+1}^n) \quad (6.22)$$

Substituting in a solution of the form $e^{-\lambda t} e^{ikx}$ yields:

$$\begin{aligned} e^{-\lambda \Delta t} - 1 &= \frac{\kappa \Delta t}{\Delta x^2} (e^{-ik\Delta x} - 2 + e^{ik\Delta x}) \\ &= \frac{\kappa \Delta t}{\Delta x^2} (2 \cos k\Delta x - 2) \\ &= -\frac{4\kappa \Delta t}{\Delta x^2} \sin^2 \frac{k\Delta x}{2} \\ e^{-\lambda \Delta t} &= 1 - \frac{4\kappa \Delta t}{\Delta x^2} \sin^2 \frac{k\Delta x}{2} \end{aligned} \quad (6.23)$$

For the scheme to be stable, the amplification factor $e^{-\lambda \Delta t}$ must be less than 1 in absolute magnitude; $|e^{-\lambda \Delta t}| \leq 1$. This leads to the stability criterion:

$$\frac{2\kappa \Delta t}{\Delta x^2} \leq 1 \quad (6.24)$$

For more than one dimension, each dimension contributes a term on the left hand side of the inequality of the same form.

Note that for the upper limit (satisfying the equality), the amplification factor is negative. This implies that λ is imaginary and so the solution is oscillatory. To ensure that all wave numbers decay without spurious oscillations, one must satisfy the criterion:

$$\frac{4\kappa \Delta t}{\Delta x^2} \leq 1 \quad (6.25)$$

In principle, at the upper limit of this inequality, a pure grid length wave ($k = \pi/\Delta x$) is instantaneously damped to zero amplitude.

This simple example is presented as an analogue for analyzing the stability of the discretisation described in this chapter. The general form for diffusion in the finite volume model is:

$$\frac{\partial}{\partial t} \phi = \frac{\Delta t}{V} \delta_x \left(\frac{\kappa A}{\Delta x} \delta_x \phi \right) \quad (6.26)$$

where V , A , κ and Δx can all be functions of x . For convenience, define:

$$\begin{aligned} \gamma_{i+\frac{1}{2}} &= \frac{\Delta t \kappa_{i+\frac{1}{2}} A_{i+\frac{1}{2}}}{V_i \Delta x_{i+\frac{1}{2}}} \\ \gamma_{i-\frac{1}{2}} &= \frac{\Delta t \kappa_{i-\frac{1}{2}} A_{i-\frac{1}{2}}}{V_i \Delta x_{i-\frac{1}{2}}} \end{aligned}$$

so that the amplification equation can be written:

$$\begin{aligned} e^{-\lambda \Delta t} - 1 &= \gamma_{i+\frac{1}{2}} (\cos k\Delta x - i \sin k\Delta x - 1) - \gamma_{i-\frac{1}{2}} (-\cos k\Delta x + i \sin k\Delta x + 1) \\ &= (\gamma_{i+\frac{1}{2}} + \gamma_{i-\frac{1}{2}}) (\cos k\Delta x - 1) + i(\gamma_{i+\frac{1}{2}} - \gamma_{i-\frac{1}{2}}) \sin k\Delta x \\ &= (\gamma_{i+\frac{1}{2}} + \gamma_{i-\frac{1}{2}}) \left(-2 \sin^2 \frac{k\Delta x}{2} \right) + i(\gamma_{i+\frac{1}{2}} - \gamma_{i-\frac{1}{2}}) 2 \sin \frac{k\Delta x}{2} \cos \frac{k\Delta x}{2} \end{aligned} \quad (6.27)$$

If γ varies across a grid-cell, the amplification factor has an imaginary component. The stability condition for each wave number is:

$$\left(1 - 2(\gamma_{i+\frac{1}{2}} + \gamma_{i-\frac{1}{2}}) \sin^2 \frac{k\Delta x}{2} \right)^2 + 4(\gamma_{i+\frac{1}{2}} - \gamma_{i-\frac{1}{2}})^2 \sin^2 \frac{k\Delta x}{2} \left(1 - \sin^2 \frac{k\Delta x}{2} \right) \leq 1 \quad (6.28)$$

The maximum value of the left hand side is at $k = \pi/\Delta x$ if γ is always positive (as it should be). The overall stability criterion is therefore:

$$2(\gamma_{i+\frac{1}{2}} + \gamma_{i-\frac{1}{2}}) = \frac{4\Delta t}{V} \overline{\left(\frac{\kappa A}{\Delta x}\right)^x} \leq 1 \quad (6.29)$$

which must be satisfied at each point in the model. For regular gridding the criteria reduces to 6.24.

The Courant number for the advection terms is equivalently:

$$\frac{\Delta t}{V} \overline{uA^x} \quad (6.30)$$

where $\overline{A^x}/V$, in both 6.29 and 6.30, takes the role of a grid length, Δx .

The stability criterion 6.29 has implications for shaving cells to represent topography. Volumes can not be allowed to become arbitrarily small without adversely requiring a smaller time-step. To be able to shave cells in an unrestricted manner would require an implicit evaluation of most terms in the model. An alternative is to limit the minimum volume of a cell to some fraction of an unshaved cell. This is done by either filling or emptying the cell with land by some small amount.

6.2.5 Comment on Accuracy

The accuracy of the finite volume method is second order in the interior. This is qualified because the areas and volumes are regular in the interior and so the scheme is equivalent to a second order, finite different rendition of the model.

The accuracy of models at solid boundaries is a long disputed issue (see Dukowicz and Dvinsky, 1992 [DD92]). At the boundaries, shaving cells is analogous to changing the grid-size. This means that the differencing is no longer centered so the accuracy must drop to first order for shaved cells at the boundaries.

6.3 Testing the finite volume approach

Two experiments, each of a very different nature, have been chosen to illustrate the potential of this finite volume approach for the treatment of topography. The experiments involve well known interactions with topography.

Topographic β effect: Steady, inviscid and linear flow of a homogeneous rotating fluid must, by the Taylor-Proudman theorem, follow contours of f/H . Variations in f/H can be due to a planetary vorticity gradient, β , or to a sloping bottom.

Flow over a Gaussian bump: The interaction of a zonal flow with tall isolated sea-mounts have been modelled extensively. The circulation resulting from stratified flow over a tall Gaussian bump in a periodic channel is modelled; solutions from the SPEM code (σ co-ordinate) and the finite volume approach are compared.

6.3.1 Topographic β

We configured the model on a β -plane with one homogeneous, flat-bottomed layer and forced by a sinusoidal wind-stress. The resulting flow is the Stommel solution; zonally asymmetric due to the anisotropic propagation of Rossby waves in the planetary potential vorticity gradient. On an f -plane, shallowing of the basin towards the north creates an equivalent PV gradient. This is modelled by shaving the bottom off the layer as a function of latitude. The resulting circulation can be made identical to the β -plane solution if the layer depth is appropriately chosen.

If the Rossby number is small, the potential vorticity of a homogeneous layer is f/H , where f is the Coriolis parameter and H is the depth of the fluid.

On a β -plane, the potential vorticity and its meridional gradient are, for a flat bottomed ocean of depth H_o :

$$Q = \frac{f_o + \beta y}{H_o} \quad ; \quad \frac{\partial Q}{\partial y} = \frac{\beta}{H_o} \quad (6.31)$$

On an f -plane with variable H , the meridional potential vorticity gradient is:

$$Q = \frac{f_o}{H(y)} \quad ; \quad \frac{\partial Q}{\partial y} = \frac{-f_o}{H(y)^2} \frac{\partial H}{\partial y} \quad (6.32)$$

Experiment		I	II
Dimension for x	km	0 to 2000	0 to 2000
Dimension for y	km	-2000 to 2000	-2000 to 2000
Grid spacing Δx	km	100	100
Grid spacing Δy	km	100	100
Time step Δt	s	3600	3600
Depth $H(y)$	m	4000	$\frac{4000}{1+1.6 \times 10^{-11}y}$
Coriolis $f(y)$	s^{-1}	$10^{-4} + 1.6 \times 10^{-11}y$	10^{-4}
Horizontal viscosity	m^2s^{-1}	5000	5000
A_H			

Table 6.1: Parameters for the two experiments to model planetary (I) and topographic (II) β effects.

To obtain the same PV and gradient in the both models, H in the latter must vary in the following way:

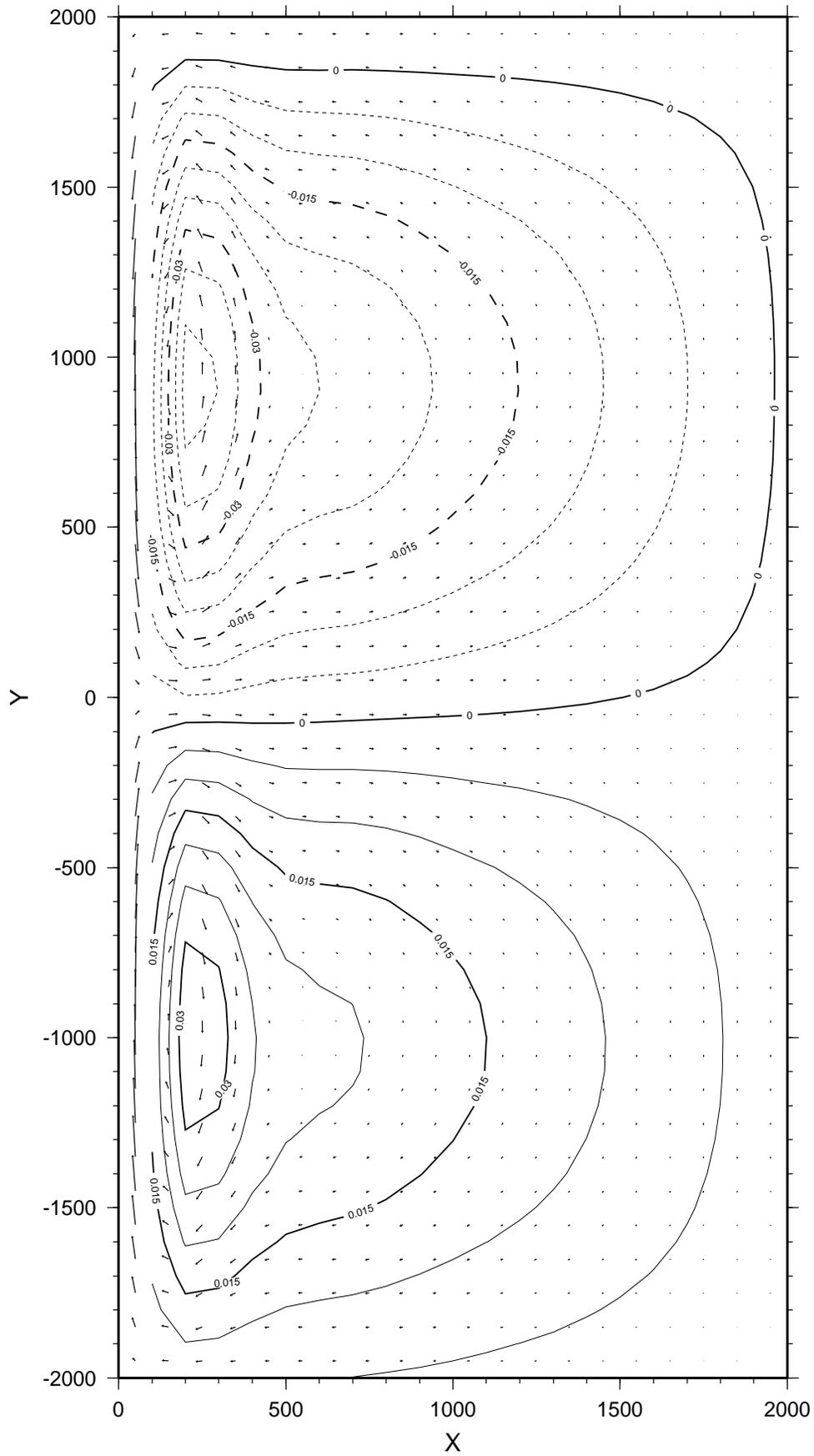
$$H(y) = \frac{H_o f_o}{f_o + \beta y} \quad (6.33)$$

The model was integrated with shaved cells in two experiments, the parameters of which are listed in table 6.1. The wind stress was the same in each experiment; $\tau^{(x)} = \tau_o \cos 2\pi y/L_y$. The maximum stress was $\tau_o = 2$ dyne cm^{-2} .

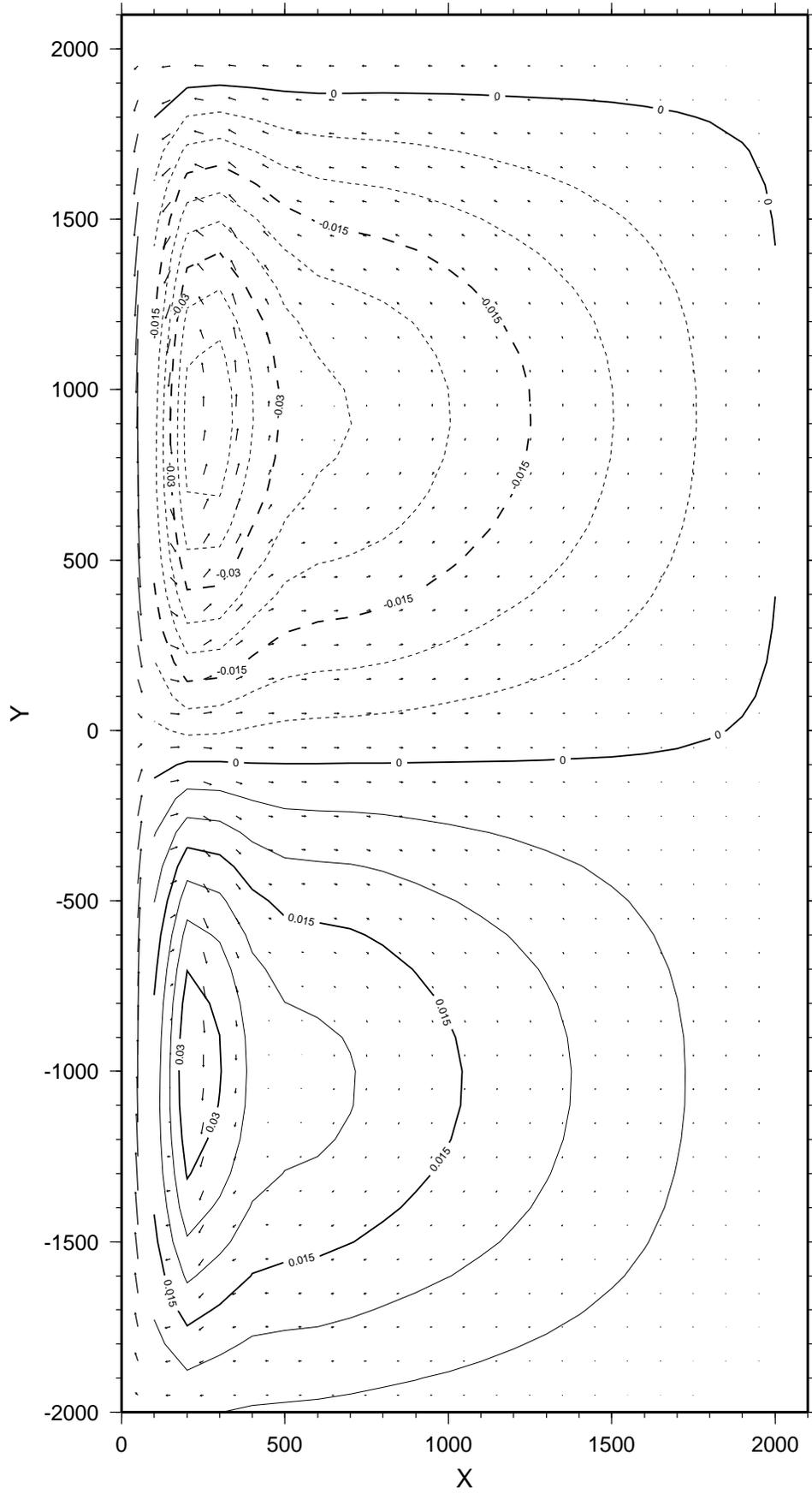
The horizontal viscosity was chosen such that the Munk layer ($\sim (A_H/\beta)^{1/3}$) was just resolved by the grid.

The slope used in experiment II meant that the depths of the basin at the north and south ends respectively were $3/4$ and $1\frac{1}{2}$ or the depth used in experiment I.

The solutions from the two integrations are shown in figures 6.4 and 6.5. The solutions are essentially linear ($R_o \ll 1$). They are very similar and thus verify that the shaved cell approach can represent smooth topographic variations.



0.10 m s⁻¹
 Figure 6.4: Pressure and velocity vectors at $t = 1$ yr for the flat bottomed β -plane integration. 90



0.10 m s^{-1}
 Figure 6.5: Pressure and velocity vectors at $t = 1 \text{ yr}$ for the sloping bottom f-plane integration.

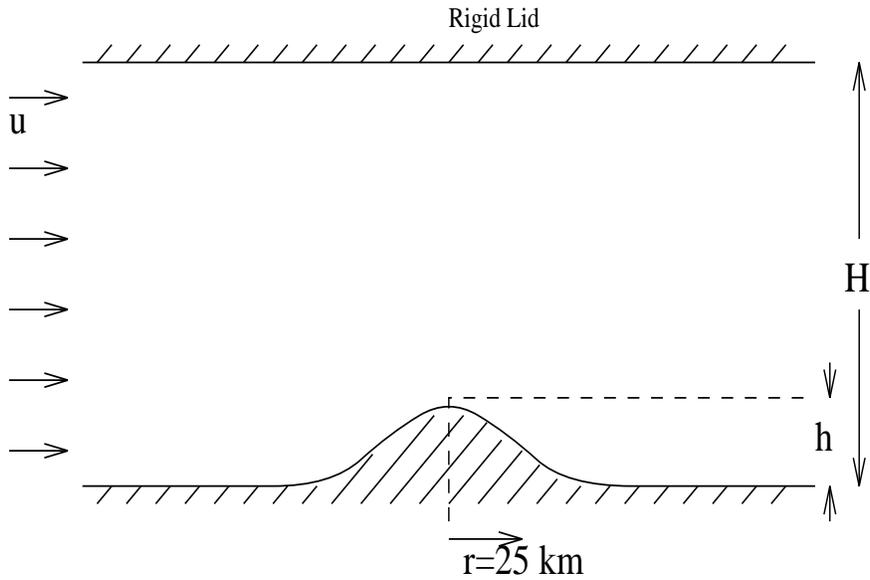


Figure 6.6: Schematic flow over an isolated topographic feature.

6.3.2 Flow over a Gaussian bump

Comparison with analytic solution

The analytical solution for quasi-geostrophic flow over topography was obtained by Huppert [Hup75]. The solution is most valid for low topography and low Rossby number flow. The analytic solution from Huppert describes the steady state and is outlined in Appendix C.

The finite volume model was set up in a $400 \text{ km} \times 300 \text{ km}$ channel with 5 km horizontal resolution and 8 levels in the vertical, each 562.5 m thick. A Gaussian bump, of height 225 m (5% of the channel depth) and e-folding scale 25 km , was placed in the centre of the channel. The bottom layer was shaved appropriately to retain the volume and cross-sectional areas of the domain.

The model was initialized with a uniform zonal flow of 5 cm s^{-1} . A cyclonic eddy is generated near the bump and advected downstream. An anti-cyclonic eddy remains over the topographic feature.

The barotropic stream function with the mean zonal flow removed is shown at day 60 in figure 6.7. The Gaussian bump is centered in the figures. For comparison, the analytic steady state solution has been subtracted from the model output and the residual shown in figure 6.8. The analytic steady state solution assumes that the cyclonic eddy has shed and been advected downstream to an infinite distance. In the numerical calculation, the domain is periodic and so the eddy re-enters the channel from upstream. The difference between the model and analytic results is dominated by the cyclonic eddy and is otherwise quite small.

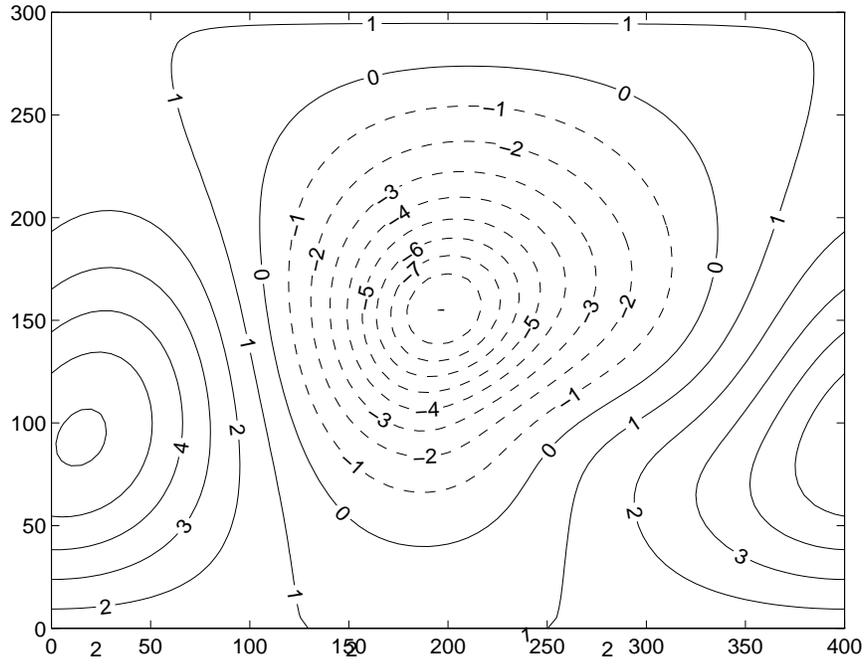


Figure 6.7: Barotropic stream function at day 60 with constant zonal flow removed.

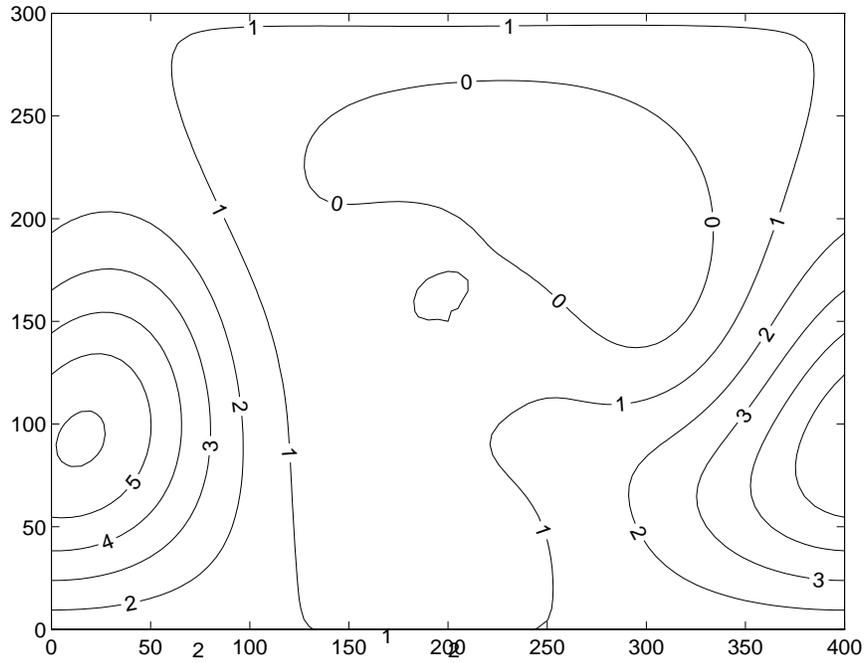


Figure 6.8: Barotropic stream function at day 60 with analytic solution removed.

		FVGCM	SPEM
Barotropic in-flow u_o	cm s ⁻¹	25	25
Zonal resolution Δx	km	5	4.1*
Meridional resolution Δy	km	5	6.1*
Nominal ocean depth H	m	4500	4500
Height of bump h	m	4050	4050
Length scale of bump L	km	25	25
Stratification $\frac{NH}{fL}$		1.5	1.5
Horizontal Viscosity $A_H \nabla_h^2$	m ² s ⁻¹	100	0
$A_4 \nabla_h^4$	m ⁴ s ⁻¹	0	5×10^{-9}
Vertical Viscosity $A_V \frac{\partial^2}{\partial z^2}$	m ² s ⁻¹	10^{-3}	0
Horizontal Diffusion $\kappa_H \nabla_h^2$	m ⁴ s ⁻¹	10	0
$\kappa_4 \nabla_h^4$	m ² s ⁻¹	0	1×10^{-9}
Vertical Diffusion $\kappa_V \frac{\partial^2}{\partial z^2}$	m ² s ⁻¹	10^{-5}	0
Time step Δt	s	600	432

Table 6.2: Parameters for the the comparison experiment to SPEM. The resolution in the SPEM integration is approximate. A stretched co-ordinate was used to increase resolution over the bump.

Comparison with SPEM code

A comparison of the finite volume code with SPEM (version 3.0) was recently made possible by Dan Goldner (MIT/WHOI Joint Program, 1995). He made available some output from an experiment he conducted using SPEM to study flow over a seamount (Fieberling).

A Gaussian bump was placed in a periodic channel of width 300 km and length 400 km. The bump had a characteristic length scale of 25 km and was centred in the channel. It rose to a height of 90% the depth of the ocean (*i.e.* to within 450 m of the surface).

The finite volume code was configured to match the SPEM integration as much as possible. SPEM uses a spectral representation in the vertical. 8 modes were used in SPEM so 8 levels of the finite volume model were chosen, equally spaced since the stratification was initially constant. SPEM is formulated using σ co-ordinates as a vertical co-ordinate. It used stretched co-ordinates in the horizontal to selectively increase resolution. A horizontal resolution of 5 km was chosen for the finite volume model as an approximate mean of that in the SPEM resolution.

SPEM used bi-harmonic dissipation, a form not yet implemented in the finite volume code at the time of this work. The bi-harmonic dissipation gave a time-scale of 35 hrs at length scale of 5 km. The horizontal mixing co-efficient was chosen to give a longer diffusive time-scale of 70 hrs, reducing the damping at longer scales.

The models were initialized with a barotropic inflow of 25 cm s⁻¹. The flow in both models is immediately deflected to the left (see figures 6.9 and 6.10) as it passes over the bump. In time, an anti-cyclonic and cyclonic eddy are formed, the latter of which is shed off the bump and advected down stream (see figures 6.11, 6.12, 6.13 and 6.14).

The effect of the topography in steering the flow is similarly represented in both models. The shedding and advection of the cyclonic eddy is somewhat slower in the finite volume code. The differences are most likely due to the differences in resolution, both vertical and horizontal. The discrepancies in dissipation may also be responsible. Bi-harmonic dissipation has not yet been implemented in the finite volume code. There was no available output from SPEM that did not use bi-harmonic viscosity.

The principle result is that the finite volume formulation can represent order one topographic variation. The difference in detail between the two integrations are almost definitely mostly due to the details of the individual models (such as different advection schemes, co-ordinate systems and resolution).

It is not clear, with out conducting a series of integrations with more direct control of SPEM, whether the accuracy of the finite volume method should be brought into question. Such a series of tests should definitely be made, but have not yet been done so. Nevertheless, it is clear that the effect of topography on the zonal flow are definitely realised by the finite volume approach.

Comparison of shaved cells vs step topography

The most important question regarding the use of shaved cells to represent topography is: How do the results compare to the conventional step-wise representation? Here, we repeat a similar experiment to the low Rossby number calculation of section 6.3.2 to answer this question.

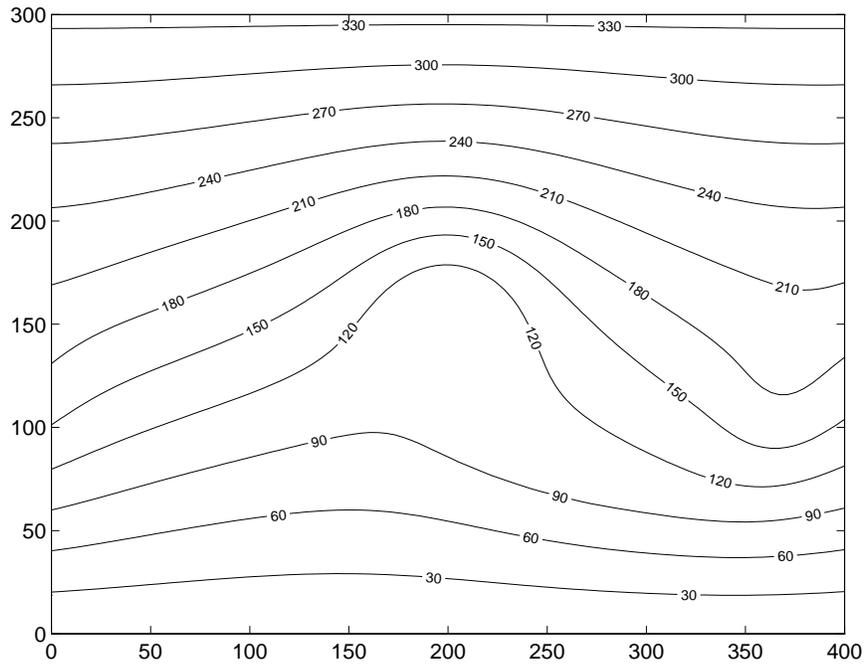


Figure 6.9: Barotropic stream function Ψ from the SPEM integration at $t = 10$ days. Contour interval = 30 Sverdrups.

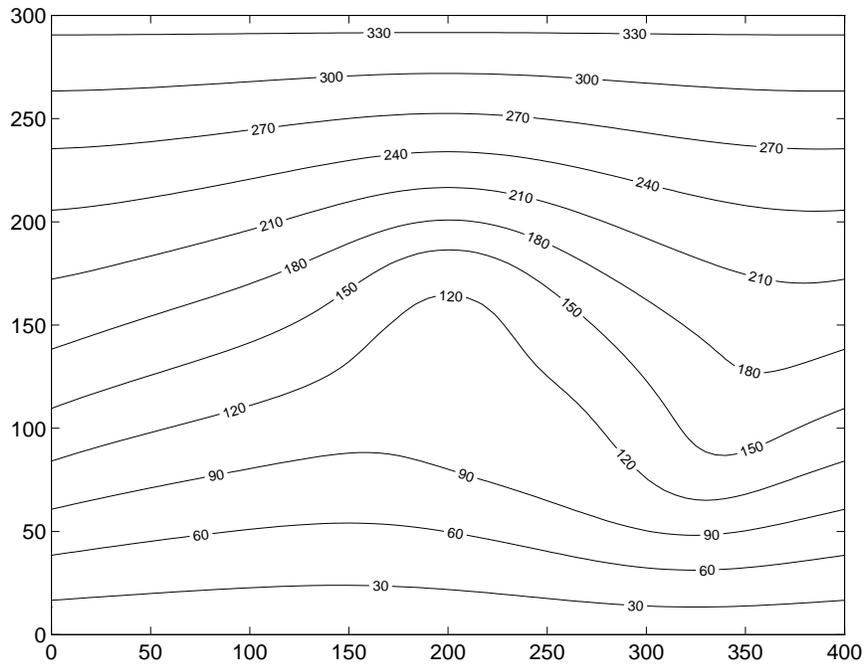


Figure 6.10: Barotropic stream function Ψ from the shaved cell integration at $t = 10$ days. Contour interval = 30 Sverdrups.

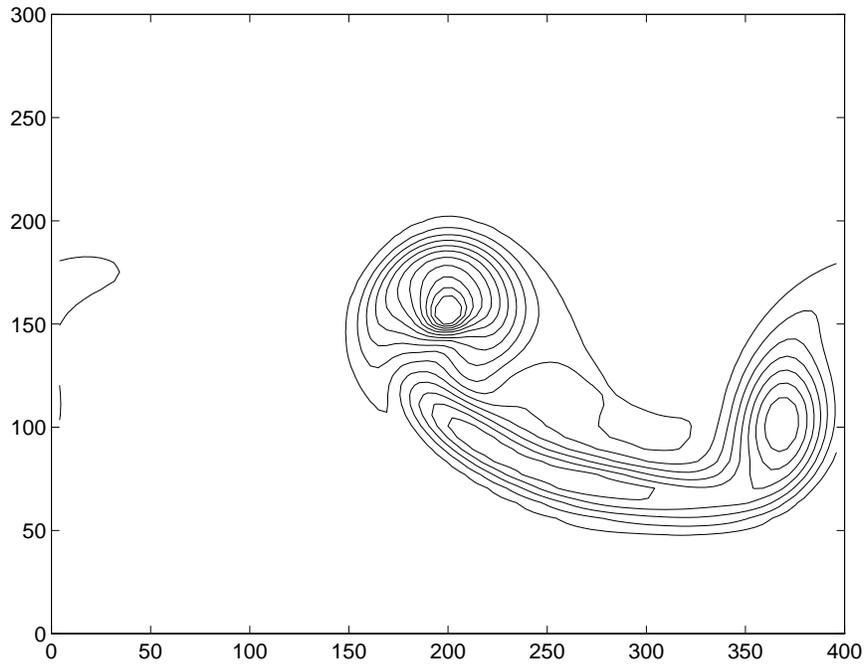


Figure 6.11: Depth Integrated Relative Vorticity in the SPEM integration, $t = 10$ days.

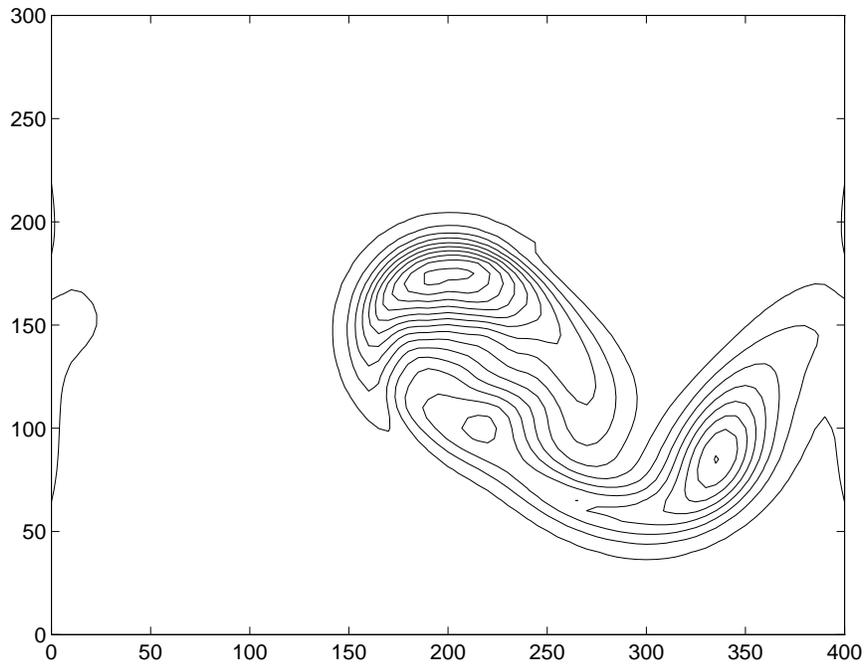


Figure 6.12: Depth Integrated Relative Vorticity in the shaved cell integration, $t = 10$ days.

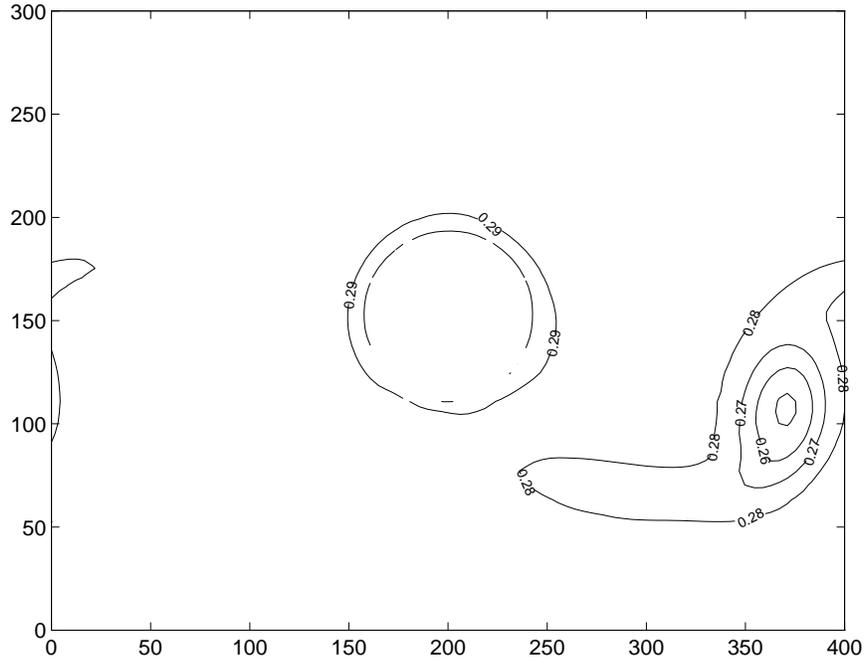


Figure 6.13: Density anomaly at $z = -4000$ m in the SPEM integration, $t = 10$ days. Contour interval = 0.01 kg m^{-3} .

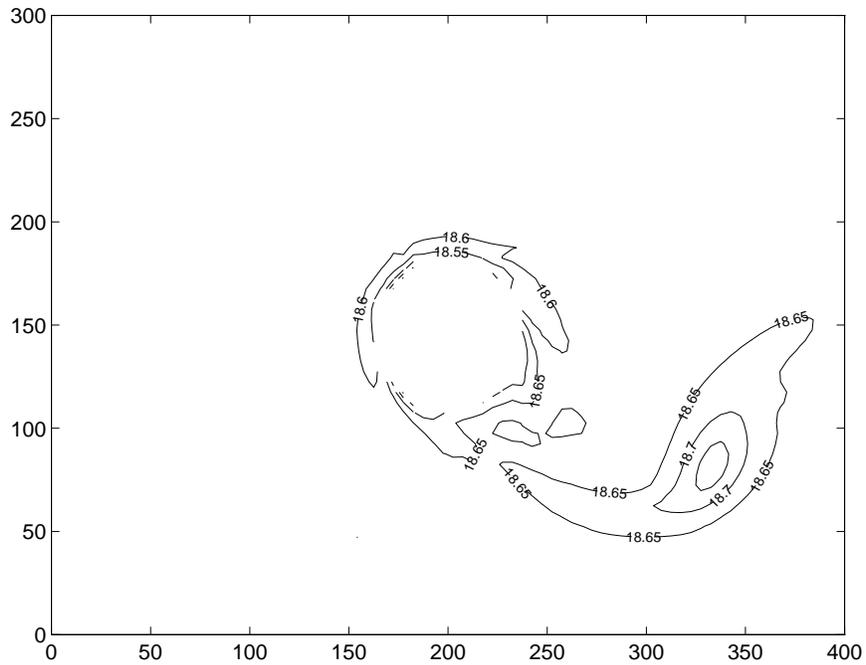


Figure 6.14: Potential Temperatures at $z = -4219$ m in the shaved cell integration, $t = 10$ days. Contour interval = $0.05 \text{ }^\circ\text{C}$.

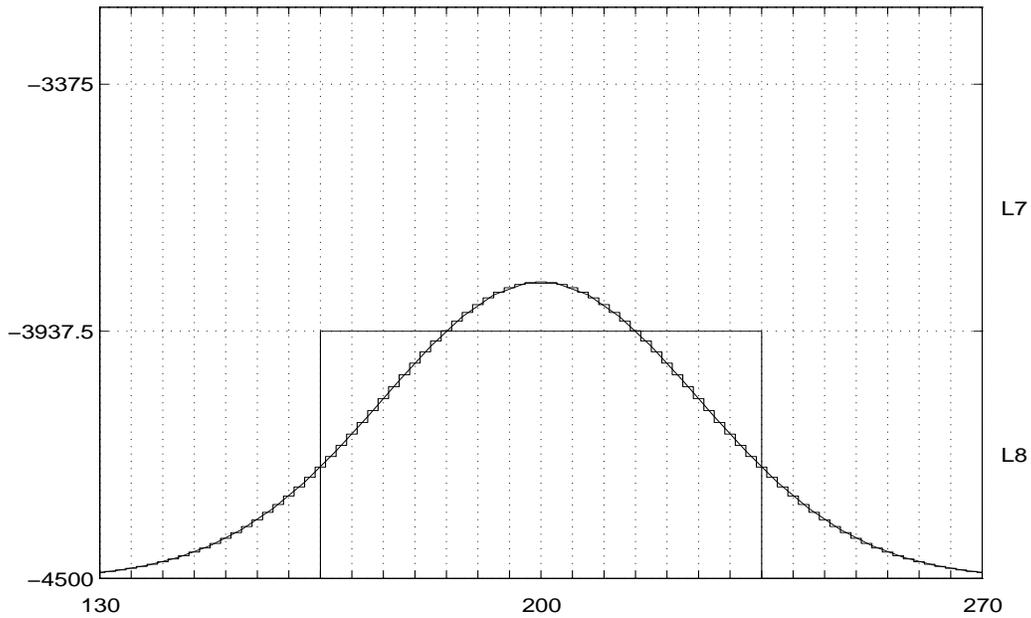


Figure 6.15: Schematic of the bottom two layers in the eight layer integrations using first “step-topography” and then “shaved-cells” to represent a Gaussian bump. The regular model cells are indicated by the dotted grid. The continuous curve is the Gaussian topography to be modelled. The large pill-box curve is the “best” representation of the Gaussian that can be managed using step-wise topography. Since the graphical presentation of areas and volumes is awkward, the curve used to calculate the volumes and areas is presented. The curve is described on a grid of three times higher resolution than that of the model. It should be apparent to the eye that the volumes and areas used in the finite volume calculation are consequently for more representative of the “true” volumes and areas than in the step-wise calculation.

The Gaussian bump now has a height of 15% of the fluid depth. Figure 6.15 shows the analytic, step-wise and finite-volume representations of the Gaussian function. The model cells are depicted by the dashed grid. The horizontal resolution of 5 km amply resolves the 25 km exponential scale of the Gaussian function. The analytic function penetrates the second deepest layer of the model. The step-wise representation of the Gaussian bump is as a pill-box, of crudely the same volume, that fills cells in the bottom layer of the model. The stair-case curve is the numerical approximation to the analytic function used for calculating the volumes and areas used in the finite volume method.

The volumes and areas are calculated as follows. The analytic function is evaluated on a fine grid of three times higher resolution and these values are used as the height of pill-boxes. The volumes of the pill-boxes are then summed up over each model cell and the result is used for the volume of the model cell. The face area of model cells is calculated by summing up the face areas of the fine-grid pill-boxes abutting a model face. As the resolution of the fine-grid is increased, the method asymptotes to an analytic method of evaluating the volumes and areas.

The models were initialised and integrated as before. Figures 6.16 and 6.17 show the potential temperature at a depth of 3656 m (corresponding to the second deepest layer) at the end of 1 day. In the shaved cell integration (6.17), the anti-cyclonic eddy is moving clockwise around the bump whilst the cyclonic eddy is beginning to be influenced by the mean zonal flow. The step-wise integration is significantly different. The anti-cyclonic eddy has not been carried clockwise around the bump, possibly because of the angularity of the feature. The strength of the eddies is approximately the same, though they are closer in separation, presumably because the feature is narrower. In short, the solution for the step-wise representation of the Gaussian bump bears minimal resemblance to the solutions of the previous sections. Indeed, the solution must presumably look more like that for flow over a cylinder, since that is what the topography of the model looks like.

In this example, the horizontal scale of topographic variation was resolved. To resolve the Gaussian bump accurately in the vertical, the conventional step-wise method would require about 30 layers in place of the bottom 1.5 layers. The finite volume method allows the model to realise the slowly varying height of the bump even though most of the variation occurs within just one equivalent model layer.

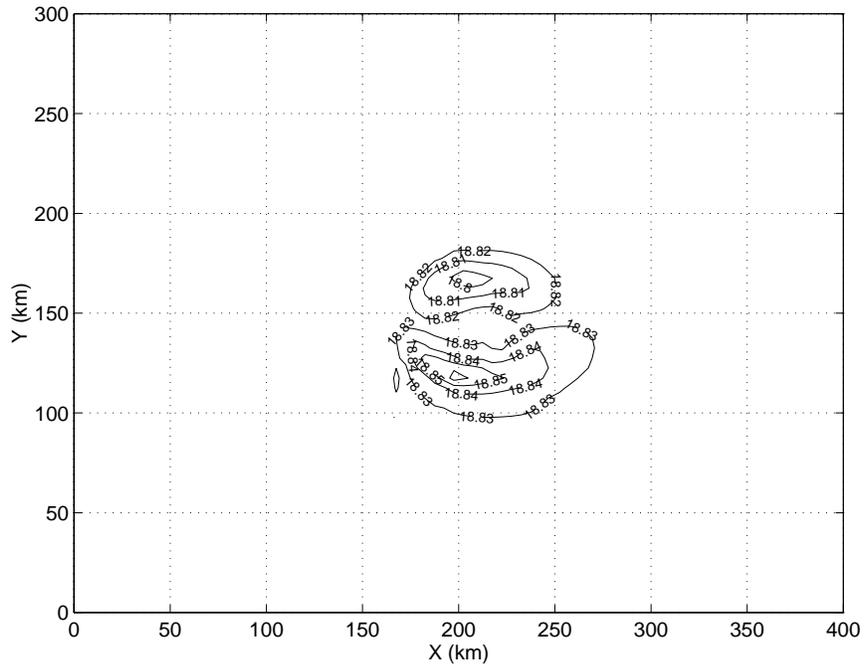


Figure 6.16: Potential temperature at $z=-3656$ m and $t=1$ day using the step-wise representation of topography.

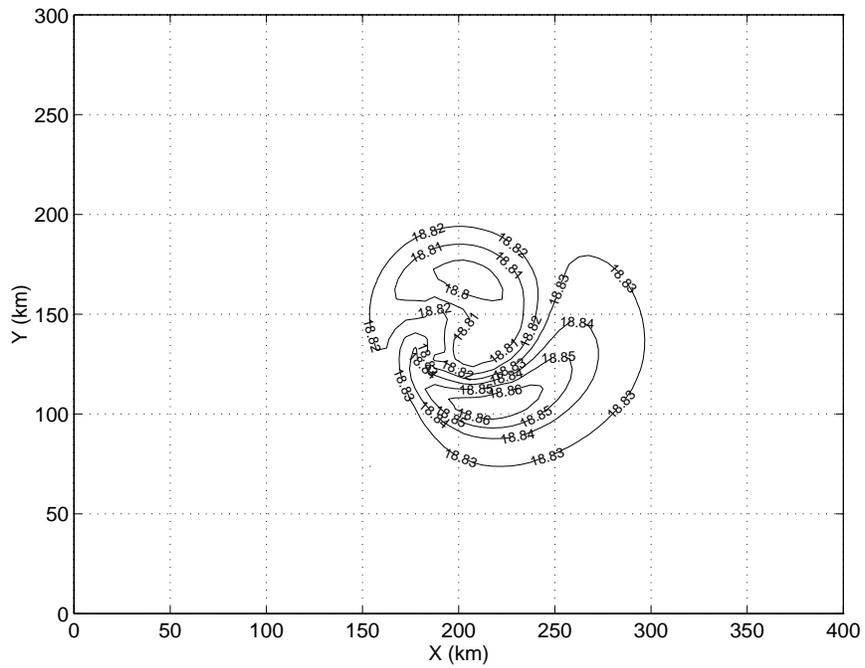


Figure 6.17: Potential temperature at $z=-3656$ m and $t=1$ day using the shaved cell representation of topography.

6.4 Conclusions

A finite volume approach to modelling the ocean has been outlined. The final form of the scheme was dictated by considerations of energy conservation. The opportunity for implementing such a technique arose because the original model was formulated on a C grid.

The motivation for developing the finite volume model was to improve the representation of topography. In particular, topography controls the circulation by modulating the passages through which basins communicate. Wave interaction with finite slopes also influences the circulation.

Two experiments illustrated the effects of topography and the potential of the scheme to model them better. The topographic β effect was succinctly illustrated. Taylor cap formation was modelled and the results compared to an almost equivalent SPEM integration. The details of the simulations differed and it is supposed, though not demonstrated, that these differences led to the different shedding time of the cyclonic eddy.

A third set of experiments compared the shaved-cell method with the conventional step-wise representation of topography. The results clearly indicate that the finite volume approach is an improvement over step-wise topography.

The accuracy of the scheme is left in question, a conclusion drawn from the comparison with SPEM code. Nevertheless, the reproduction of analytic solutions and the clear improvement over step-wise topography suggest that the shaved cell method could substantially improve the representation of topographic effects in existing height coordinate models.

Chapter 7

Concluding remarks

Of the ocean models in existence, most can be grouped into one of two categories; large scale circulation models and small scale process study models.

Large scale circulation models are designed for studying the slow large scale circulation patterns of gyres and the meridional overturning. They are typically used:

- to study the role of wind and thermal forcing in determining these large scale patterns.
- to study interactions with the atmosphere.
- to understand the role of the ocean in the climate system.
- in a diagnostic role with observations to estimate the present state of the ocean circulation.

Process study models, on the other hand, are used to study the dynamical behaviour of the oceans on regional and smaller scales which often involve resolving relatively fast processes (such as gravity waves). These include studies of the convective overturning process, the nature of dynamical instabilities, the formation of boundary layers and subsequent interaction with the atmospheric boundary layer, coastal dynamics and forecasting on a regional basis.

Large scale models and process study models differ in horizontal resolution and often also differ in physics. The large scale studies are inevitably carried out at low resolution; the Rossby radius of deformation is typically 20-60 km whilst affordable horizontal resolutions tend to be around 50 km upwards. Process studies are normally carried out at resolutions that at least resolve the geostrophic eddy scale and, if necessary, resolve the convective plume scale, depending on the process being studied.

None of the existing numerical models are well suited to work at both ends of the spectrum. The particular physics and numerical formulation are decided by the scale at which the model will be applied. A large scale model is not well suited for use at the small scale and vice versa. This means that the models are also potentially ill-suited for use in the middle ground, *i.e.* at the resolutions that are comparable to the scale of transition for validity of the physical approximations. In particular, the validity of the primitive equations at the mesoscale and smaller is questionable.

There are two main reasons for this “spectral gap” in numerical models:

- The smaller scale physics is computationally more expensive to resolve. The non-hydrostatic physics requires the inversion of a three-dimensional elliptic operator whilst hydrostatic physics inverts a two dimensional problem. For this reason, a non-hydrostatic model has not been used to study the large scale circulation.
- The different extremes in horizontal resolution dictate that different grids be used for large and small scale models. As was explained in chapter 4, coarse resolution models tend to be built on an Arakawa ‘B’ grid and high resolution models (and non-hydrostatic models) are normally built on an Arakawa ‘C’ grid. The application of a ‘C’ grid model at low resolution, or a ‘B’ grid model at very high resolution are prone to grid-scale problems that normally inhibit the forward integration of the models.

Part of the motivation behind building a new ocean circulation model was to build one applicable to both ends of spectrum of spatial scales, and thus build one applicable in the middle ground. Such a model then allows us to evaluate when certain physical approximations become invalid. By no means do we expect the global circulation to be substantially affected by non-hydrostatic effects, but regional calculations, for example at the Equator, will quite possibly be dependant on the inclusion of quasi- or non-hydrostatic physics. The advantage of such a versatile model

is that the transition between small, regional and global scale calculations is not complicated by the use of different codes.

The non-hydrostatic model, described in chapter 3, can be integrated forward at low resolution on a global scale. This is only affordable because of the nature of the pre-conditioner used in the conjugate gradient algorithm. Though the results are not shown here, the non-hydrostatic physics is found to have an insignificant effect on the large scale circulation. This is as expected from conventional understanding and scaling analyses. The question still remains, at what scale does non-hydrostatic physics begin to play a role? This is not just concerned with the role of vertical inertial effects in the ocean but also with the role of the horizontal Coriolis terms at low latitudes.

7.1 The C_d scheme

An example of the problems arising from using the Arakawa ‘C’ grid at low resolution was given at the end of chapter 3. The nature of the problem is connected with the spatial averaging of the Coriolis term on a ‘C’ grid. This is a well documented phenomena and is described and explained in chapter 4.

A new scheme is described and tested in chapter 5 that alleviates this problem. The C_d scheme allows the ‘C’ grid model to be integrated forward without any spurious noise problems. It has little computational overhead and involves only the introduction of two new horizontal velocity variables, carried on a ‘D’ grid, which are step forward in parallel to the C grid model. These new variables are used in the evaluation of the Coriolis term for the C grid. We view the scheme as a tool used to improve the representation of the Coriolis term without directly modifying the rest of the model.

The model code is consequently now applicable at both high and low resolutions. The C_d scheme is, however, switched off at high resolutions since it is not needed. As it stands, the model has only gone part way towards the goal of applicability at high and low resolutions since the use of the C_d scheme at low resolutions means that it is not quite the same model as used at high resolution.

7.2 Shaved cells

Strictly speaking, the opportunity to explore the use of shaved cells to represent topography is a result of the model being formulated on an Arakawa ‘C’ grid. The finite volume approach naturally results in a ‘C’ grid arrangement of model variables.

Conventional height co-ordinate models represent topography by fitting the bathymetry to the discrete depths of the model layers, resulting in “step-wise” topography. As an alternative, the finite volume approach allows model cells to be shaved to fit the topography.

The model is re-formulated in chapter 6 using the finite volume approach. In practice, the resulting algebraic equations on a regular grid are essentially the same as for the finite difference model described in chapter 3. The resulting discretisation of the finite volume model depends on a restriction imposed on the nature of the elliptic problem.

A pre-conditioned conjugate gradient algorithm is used to solve the elliptic problem. The algorithm is applicable because the matrix is symmetric. It was desired that this property be retained for the purposes of efficiency. This meant that the finite volume approach could not be applied in a completely general manner to the momentum equations. Nevertheless, the effects of topography are captured in the discrete form of the continuity and tracer equations. The criteria for the discretisation of the momentum equations is based primarily on consideration of energy conservation.

Two distinct examples of oceanographic interest are used to demonstrate the potential of using shaved cells to represent topography. The topographic β effect is modelled and the solution compared to a Stommel solution obtained on a β -plane. The solutions are indistinguishable.

The interaction of flow with an isolated topographic feature is modelled by placing a Gaussian bump in a channel with a barotropic zonal flow. A low Rossby number solution is compared to an analytic solution and shown to be favourably accurate. Comparison of a high Rossby number solution with one obtained using the SPEM code demonstrates that the shaved cell representation allows the model to realise the effects of extreme topography. Two equivalent integrations, one using a step-wise representation of the topographic feature, the other using the shaved cell method, are compared. The shaved cell method clearly represents the topography more accurately than the conventional step-wise method.

The method is thus shown to be able to resolve both shallow and deep changes in topography, even with a small number of levels in the vertical. The same will apply in the horizontal, meaning that coastlines can be represented using the same shaved cell method. One consequence for global ocean modelling is that despite the usually course

horizontal resolution, certain features in the coast-lines may still be realised by the circulation. For example, coastlines are conventionally represented by meridional or zonal barriers whilst the shaved cell method will allow a coast to be diagonal across meridians. Just as the height of topography is conventionally made to fit the model depths, the positions of coasts are moved to fit the model grid. This implies that passages are either widened or narrowed. The shaved cell method does not have this problem. Indeed, the shaved cell method allows a passage to be narrower than the grid-size. Clearly, the dynamic processes involved in flow through passages, such as hydraulic control, are not resolved but the ability to transport properties through the passage is retained. Whether the transports are accurate is not clear at this point.

7.3 Future development of model

The model is currently being used for a wide range of applications, from studies in convection, though studies of mixed layer dynamics, up to regional simulations and global simulations. Much of this work is possible because of the parallel implementation of the model. The parallel design of the non-hydrostatic model is satisfactorily optimal. The two-dimensional solver does not take advantage of the parallel architecture due to the choice of mapping to the processors.

A move to a free surface formulation is planned that we anticipate will improve the performance of the two-dimensional solver. Dukowicz and Smith, 1995 [DS94], point out that elimination of the rigid-lid approximation has several advantages; among others it computes the surface height directly, allowing comparison with and assimilation of altimetry data and it improves the accuracy of the representation of long Rossby waves.

This issue is raised here because this model has a significant advantage over some others in the use of a free surface. The free surface formulation re-introduces the external inertia-gravity mode which is filtered out by the rigid-lid approximation. The external mode has a phase and group speed of \sqrt{gH} , as shown in section 2.3.1. The important point is that the deformation radius, $\sqrt{gH/f^2}$, is of the order of 1000 km. This means that the wave resolution is inevitably very high. In this limit, a ‘B’ grid model is severely prone to a checker board mode in the pressure field. In contrast, the ‘C’ grid model is very well suited for resolving the external gravity wave motion and does so accurately.

7.4 Future directions in ocean modelling

Historically, the low resolution restrictions imposed by the computer technology has lead to an emphasis in the development of sub-grid scale parameterizations. Even with the recent improvements in computer architectures, there is still a need for better parameterizations of all sub-grid scale processes. This will always be the case until computer technology advances far enough to allow large scale models to resolve the three dimensional turbulence scales.

The need for future model development should also not be underestimated. The need for more versatile models has been outlined. The general circulation model described here has been designed to be applicable at many scales. However, there is a further degree of versatility that has yet to be made accessible which is the ability to model a wide range of time-scales.

Much current and future climate research explores the natural variability and the predictability of the climate system on time-scales from seasons upwards to centuries. There is much interest in devising methods to find and characterize steady states of the thermohaline circulation, and the stability of these states to perturbations. The above all involve extended prognostic integrations over periods long relative to the dynamical time scales of the individual sub-systems.

In ocean models based upon the primitive equations, there remains a practical restriction on the time step required for making long integrations. The restriction is due to the natural frequencies of the physics explicitly carried in the model. One avenue of investigation has been to further approximate the physics, filtering out intermediate time scale processes (eg. gravity waves) by careful *scaling* of the equations of motion. This has led to the development of a whole class of intermediate models (eg. Balance Equations, Semi-geostrophic models, etc) that have proved useful and certainly will be for some time to come. However, these intermediate models can be difficult to solve and are often limited in their application to other than idealized studies. One example of a particularly successful intermediate model is the Planetary Geostrophic Equations (see Maier-Reimer et al., 1993 [MRMH93]). This model has been integrated for thousands of years in various studies of variability of the thermohaline circulation. However, the model does have approximated physics and thus excludes the possibility of some processes.

An alternative to changing the continuous equations, is to integrate the unapproximated equations implicitly. This technique has advantages since if the resolution is such that an otherwise implicit process is resolved, then it

will be accurately represented. Implicit evaluation of physics often leads, however, to a more complicated numerical algorithm, but interestingly one which often resembles the *filtered* models.

Implicit techniques are already frequently used to represent physics in ocean modeling. The simplest, common example is the implicit evaluation of the Coriolis term in the models based on the Arakawa B grid, such as the GFDL MOM code. This allows a time step longer than the half period of an inertial oscillation and in this case, the inertial oscillation is damped.

Another example is that of filtering the external gravity wave. Here, an explicit free surface would typically require a time-step of around 10 minutes. The rigid lid approximation filters this mode and makes the effective wave speed infinite and produces an elliptic equation to be solved for the surface pressure at each longer time-step of the model. If the free surface is treated implicitly then an elliptic problem again arises but slows the waves down to whatever the resolution can sustain. For very long time steps, the rigid lid and implicit free surface models actually converge in their formulation which is consistent with, and in some ways is a validation of, the approximation at longer time scales.

There is, in fact, a great deal of similarity, from an algorithmic point of view, between filtered models and implicit models based upon the HPEs. There is, as a result, some comparison in computational efficiency between the two classes of model. Combined with the simple practical need to integrate the HPEs with long time steps, a return to implicit methods is possible and certainly worth pursuing.

A certain proportion of the numerical knowledge needed to make models fully implicit already exists. Ocean modellers have been taking advantage of some of these methods but as of yet, no physically complete model has been written on an implicit basis. Atmospheric GCMs generally incorporate the gravity waves implicitly. The approach used there would obviously be a sound starting point. Development of numerical methods concentrating on shifting ocean models onto an implicit footing would be a significant improvement on our current ability to study the long time scales mentioned before.

The objective of making ocean models fully or semi-implicit has several immediate applications:

- i) Ocean models with very long time steps essentially become steady state solvers. The steady state is interesting in its own right but is inevitably also a common starting point for many other studies such as sensitivity studies.
- ii) Because the complete physics is still incorporated within the model, the same model used to find the initializing steady state can be used in process studies, dependent upon the *fast* physics.
- iii) The importance of the *fast* physics in climate variability and stability can be ascertained by using the same model in its various degrees of implicitness.

As already mentioned, implicit physics often leads to a more complicated algebraic formulations of the model. The solution techniques currently employed already outperform the equivalent explicit models. Indeed, the increase in computational efficiency gained by use of modern techniques (multi-grid, pre-conditioned conjugate gradient) is in fact comparable to that gained by recent innovations in computer hardware. Nevertheless, development and application of these techniques in ocean modeling is only in its infancy and as of yet has only recently begun to take advantage of the new parallel computer technologies.

The work of this thesis represents only a small step towards the goal of developing a fully versatile model. The applicability of the model to many spatial scales is only a beginning. With no changes to the kernel necessary, the code accommodates both periodic domains, curvi-linear coordinate systems, hydrostatic, quasi- and non-hydrostatic physics. Perhaps the most beneficial innovation is the use of the finite volume formulation that not only allows very easy inclusion of any arbitrary topography but further naturally handles the boundary conditions. The final code is one that can be applied in a wide range of oceanic studies.

Appendix A

Derivation of the Navier-Stokes Equations of Oceanic Motion

Here, the Navier-Stokes equations that govern fluid flow will be derived in an inviscid context. The accompanying thermodynamics will also be described to yield the fundamental equations that describe oceanic motions on a rotating planet. The derivation is concise but accurate.

Excepting for the discussion on thermodynamics, the derivation is made in an Eulerian frame rather than the conventional Lagrangian frame. The Eulerian frame is concerned with the rates of changes of quantities per unit volume, which is a convenient property when formulating discrete, numerical models. The Lagrangian frame is more conventional because the total derivative is the rate of change following a particle. It deals with the change in concentration of quantities, or mass of quantity per unit mass of fluid. The Navier-Stokes equations are summarised in section 2.1.

A.1 Conservation of Mass

A fundamental principle in fluid mechanics is the conservation of mass. For a fluid of density ρ (density is mass per unit volume), the rate of change of mass in a small volume, $\frac{\partial}{\partial t} \iiint \rho dV$, must equal the accumulation of mass flux into the volume, $\iint \rho \mathbf{u} \cdot \hat{\mathbf{n}} dS$. In the limit of infinitesimal volume, the statement of conservation of mass becomes:

$$\frac{\partial}{\partial t} \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (\text{A.1})$$

The mass flux, $\rho \mathbf{u}$, that appears in the Eulerian form, is in fact the momentum density (momentum per unit volume). This quantity will explicitly appear in the momentum equations and scalar conservation equations.

Although mass conservation is the most basic equation governing fluid flow, it is rarely used in the form of A.1 to model the ocean due to ill-conditioning of the system. The cause of this lies partly in the fact that typical density variations are small compared to the mean value. Further, density variations due solely to the compressibility of the fluid are even smaller. This means that the flow is approximately non-divergent and this condition will later replace the continuity equation.

Conservation of mass can be written:

$$\frac{\partial}{\partial t} \rho + \nabla \cdot (\rho \mathbf{u}) = \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{u} = 0 \quad (\text{A.2})$$

where $\frac{D}{Dt}$ is the Lagrangian or total derivative; the rate of change following the motion. For a non-divergent fluid, $\nabla \cdot \mathbf{u} = 0$, the density is conserved in the Lagrangian sense.

The non-divergence condition itself, $\nabla \cdot \mathbf{u} = 0$, is referred to as the continuity equation, the reason being that non-divergence is a statement of continuity of volume. These issues will all be discussed later. In the mean time, equation A.1 will be retained.

A useful relationship that might aid the reader for the rest of this derivation, transforms the Lagrangian or total derivative to a true flux form Eulerian derivative by making use of the continuity equation. For any scalar field, ϕ :

$$\rho \frac{D\phi}{Dt} = \rho \left(\frac{\partial \phi}{\partial t} + \mathbf{u} \cdot \nabla \phi \right) = \frac{\partial (\rho \phi)}{\partial t} + \nabla \cdot (\rho \phi \mathbf{u}) \quad (\text{A.3})$$

The total derivative is the rate of change following the parcel of fluid whilst the last form expresses the rate of change for a fixed volume.

A distinct advantage of the flux form is that it can be immediately integrated over the volume of the fluid to provide a statement of global conservation. For instance, in the absence of net internal sources and sinks, the total amount of some property is governed by:

$$\frac{\partial}{\partial t} \iiint_V \rho \phi \, dV = \iint_S \rho \phi \mathbf{u} \cdot \hat{\mathbf{n}} \, dS \quad (\text{A.4})$$

where ϕ can be any internally conserved scalar concentration or unity. The special case of $\phi = 1$ indicates that the total mass of the fluid can be changed only by an influx of matter through the interface enclosing the volume. For the ocean, solid boundaries permit no normal mass flux whilst precipitation and evaporation act to add or detract water from the system.

A.2 Conservation of momentum

Newton's laws of motion state that linear momentum is conserved. For a fluid of density ρ , the relevant quantity is momentum density $\rho \mathbf{u}$ or momentum (Newton seconds, $N \, s$) per unit volume (m^3). For an infinitesimal volume $dV = dx \, dy \, dz$, the total rate of change of a component of momentum is the sum of the local rate of change (within the volume) and the convergence of momentum fluxes (arising from differing rates of advection into the volume). It will equal the net force acting in each direction:

$$\frac{\partial}{\partial t} \rho \mathbf{u} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) = \sum_i \mathbf{f}_i \quad (\text{A.5})$$

The forces acting on a fluid parcel are the pressure gradient force, the force due to the gravitation of the earth and the Coriolis and centripetal forces (resulting from being in a rotating frame of reference).

The first of these is an internal force. Pressure (force per unit area) acts on all parts of the fluid pushing equally in each direction. Variations in pressure accelerate the fluid. The force acting on a volume $dV = dx \, dy \, dz$ is the change in pressure $\frac{\partial p}{\partial x} dx$ acting across the volume through the area $dA = dy \, dz$. The force per unit volume is thus $\frac{\partial p}{\partial x}$ or:

$$\mathbf{f}_p = -\nabla p \quad (\text{A.6})$$

The gravitational potential is a function of position only and the force acting on a unit volume is given by $\rho \mathbf{g}$ where $\mathbf{g} = \nabla(GM_e/r)$. Here, G is the gravitational constant, M_e is the gravitating mass of the earth and r is the distance from the centre of the earth.

The Coriolis force is a *fictitious* force but is very important in shaping the dynamics of the ocean. Linear momentum conservation applies in an inertial frame of reference. The rate of change of some quantity, \mathbf{A} , observed in the inertial frame expressed in terms of the rate measured in the rotating frame is given by the relation:

$$\frac{D}{Dt} \mathbf{A}_i = \frac{D}{Dt} \mathbf{A}_r + \boldsymbol{\Omega} \wedge \mathbf{A}_r \quad (\text{A.7})$$

where the the subscripts i and r indicate the inertial and rotating frames. The \wedge is the vector cross product. $\boldsymbol{\Omega}$ is the rotation vector of the earth.

The velocity in the inertial frame is the velocity in the rotating frame of reference with a part due to the rotation of the co-ordinate system:

$$\mathbf{u}_i = \mathbf{u}_r + \boldsymbol{\Omega} \wedge \mathbf{x} \quad (\text{A.8})$$

Using this relationship and noting that the total derivative of the position vector is the velocity in the inertial frame, $\mathbf{u}_i = \frac{D}{Dt} \mathbf{x}$, gives:

$$\frac{D}{Dt} \mathbf{u}_i = \frac{D}{Dt} \mathbf{u}_r + 2\boldsymbol{\Omega} \wedge \mathbf{u}_r + \boldsymbol{\Omega} \wedge \boldsymbol{\Omega} \wedge \mathbf{x} \quad (\text{A.9})$$

The middle term is the *Coriolis* acceleration and the last term is the *centripetal* acceleration. The latter is only a function of position and the equivalent force can be written as the gradient of a potential and is usually combined with the gravitational potential to yield an effective potential or gravity, so that:

$$\mathbf{f}_c = -2\boldsymbol{\Omega} \wedge \mathbf{u} \quad (\text{A.10})$$

$$\mathbf{f}_g = -\rho \nabla \Phi \quad ; \quad \Phi = -\frac{GM_e}{r} - \frac{\Omega^2 r^2}{2} \cos^2 \phi \quad (\text{A.11})$$

The unapproximated form for the momentum equations is then:

$$\frac{\partial}{\partial t} \rho \mathbf{u} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) + 2\Omega \wedge \rho \mathbf{u} + \rho \nabla \Phi + \nabla p = 0 \quad (\text{A.12})$$

which can be written in Lagrangian form as:

$$\rho \left(\frac{D\mathbf{u}}{Dt} + 2\Omega \wedge \mathbf{u} + \nabla \Phi \right) + \nabla p = 0 \quad (\text{A.13})$$

The kinetic energy density (kinetic energy per unit volume) is $K = \frac{1}{2} \rho \mathbf{u} \cdot \mathbf{u}$. The kinetic energy equation can be obtained by taking the scalar product of velocity with the momentum equation [A.12](#).

$$\frac{\partial K}{\partial t} + \nabla \cdot \mathbf{F}_u = -\rho \mathbf{u} \cdot \nabla \Phi + p \nabla \cdot \mathbf{u} \quad (\text{A.14})$$

The flux on the left hand side is the mechanical energy flux

$$\mathbf{F}_u = K \mathbf{u} + p \mathbf{u} \quad (\text{A.15})$$

comprised of an advective flux of kinetic energy density and an energy flux due to pressure forces. The source terms in [A.14](#) represent conversions to other forms of energy. $-\rho \mathbf{u} \cdot \nabla \Phi$ is the conversion between kinetic and gravitational potential energy. $p \nabla \cdot \mathbf{u}$ is the work done by expansion of the fluid as it converts internal into kinetic energy.

A gravitational potential energy equation can be written:

$$\rho \mathbf{u} \cdot \nabla \Phi = \frac{\partial(\rho \Phi)}{\partial t} + \nabla \cdot (\rho \Phi \mathbf{u}) - \rho \frac{\partial \Phi}{\partial t} \quad (\text{A.16})$$

where the temporal variations in gravitational potential, $\frac{\partial \Phi}{\partial t}$, are due only to tidal effects.

The total, kinetic and gravitational potential, energy equation in the absence of tides is then:

$$\frac{\partial}{\partial t} (K + \rho \Phi) + \nabla \cdot (\mathbf{F}_u + \rho \Phi \mathbf{u}) = p \nabla \cdot \mathbf{u} \quad (\text{A.17})$$

where the only source is the conversion from internal energy.

A.3 Conservation of salt

Mass conservation also applies to the constituents of sea water. The total mass of salt in the ocean is essentially fixed over periods shorter than geological time scales, even though the local and mean salinity may vary in time and space. Salinity, S , is the concentration of salt or mass of salt per unit mass of water. The salt density or mass of salt per unit volume is ρS . For an infinitesimal volume, the rate of change of salt density must equal the accumulation of salt mass flux, $\rho S \mathbf{u}$.

$$\frac{\partial}{\partial t} \rho S + \nabla \cdot (\rho S \mathbf{u}) = 0 \quad (\text{A.18})$$

Making use of the continuity equation [A.1](#), the salinity equation (referring to the Lagrangian form) is easily obtained:

$$\frac{\partial}{\partial t} \rho S + \nabla \cdot (\rho S \mathbf{u}) = \rho \frac{DS}{Dt} + S \left\{ \frac{\partial}{\partial t} \rho + \nabla \cdot (\rho \mathbf{u}) \right\} = 0 \quad \Rightarrow \quad \frac{DS}{Dt} = 0 \quad (\text{A.19})$$

These two equations apply equally well to any scalar quantity that has no sources and is itself a concentration.

A.4 Thermodynamics (Continuity of heat, conservation of energy and potential temperature)

This section is based on the thermodynamics chapter for ‘‘Principles of Ocean Physics’’ by Apel [[Ape87](#)].

The heat equation describes changes in heat content per unit volume, ρq :

$$\frac{\partial}{\partial t} \rho q + \nabla \cdot (\rho q \mathbf{u}) = \nabla \cdot \mathbf{F}_q + \mathcal{H} \quad (\text{A.20})$$

where q is the heat content per unit mass, \mathbf{F}_q is a non-advective heat flux (normally radiation heat flux) and \mathcal{H} is other heat sources (eq. thermal vents, body heating).

The first law of thermodynamics is a statement of conservation of the internal energy. In particular, it deals with the change in internal energy resulting from a change in heat content ($\frac{Dq}{Dt}$) and the differential work done on the system ($-p \frac{D}{Dt}(\frac{1}{\rho})$). Written in Lagrangian form, the first law reads:

$$\frac{De}{Dt} = \frac{Dq}{Dt} - p \frac{D}{Dt}(\frac{1}{\rho}) + \mu \frac{DS}{Dt} + L \quad (\text{A.21})$$

where e is the internal energy per unit mass, μ is the chemical potential of dissolved salts and L is the heat of transformation.

The term representing the mechanical work done on the system can be written

$$p \frac{D}{Dt}(\frac{1}{\rho}) = \frac{p}{\rho} \nabla \cdot \mathbf{u} \quad (\text{A.22})$$

using the continuity equation.

The total internal, kinetic and gravitational potential energy equation is obtained by adding equation A.21 to equation A.17:

$$\frac{\partial}{\partial t} (K + \rho \Phi + \rho e) + \nabla \cdot (\mathbf{F}_u + (\rho \Phi + \rho e) \mathbf{u} + \mathbf{F}_q) = \mathcal{H} + \rho L \quad (\text{A.23})$$

where conservation of salt has been assumed.

Neither the heat content nor internal energy are practical quantities since they can not be directly observed. The observable thermodynamic variable is temperature, T , defined as the change in enthalpy per unit mass as the entropy varies at constant pressure and salinity:

$$T \equiv \left. \frac{\partial h}{\partial \eta} \right|_{p,S} \quad (\text{A.24})$$

An equation of state connects the temperature to the entropy η , the pressure p and salinity S ; $T = T(\eta, p, S)$. Variations in temperature can then be related to changes in entropy, pressure and salinity:

$$\frac{DT}{Dt} = \left. \frac{\partial T}{\partial \eta} \right|_{p,S} \frac{D\eta}{Dt} + \left. \frac{\partial T}{\partial p} \right|_{\eta,S} \frac{Dp}{Dt} + \left. \frac{\partial T}{\partial S} \right|_{\eta,p} \frac{DS}{Dt} \quad (\text{A.25})$$

Before the temperature equation can be used, the first term must be made measurable since entropy is not an observable quantity. At this stage the second law of thermodynamics can be introduced which states that *the change in entropy per unit mass is greater than the heat added divided by the temperature*:

$$\frac{D\eta}{Dt} \geq \frac{1}{T} \frac{Dq}{Dt} \quad (\text{A.26})$$

The equality strictly only holds for reversible processes but in practice, so long as the rates of change are slow so that the system is close to equilibrium, the equality can be assumed.

The specific heat capacity at constant pressure is defined $C_{p,S} = \left. \frac{\partial q}{\partial T} \right|_{p,S}$ and can be recast using the second law as:

$$C_{p,S} = \left. \frac{\partial q}{\partial T} \right|_{p,S} = T \left. \frac{\partial \eta}{\partial T} \right|_{p,S} \quad (\text{A.27})$$

The first term in the temperature equation can therefore be rewritten:

$$\left. \frac{\partial T}{\partial \eta} \right|_{p,S} \frac{D\eta}{Dt} = \frac{T}{C_{p,S}} \frac{D\eta}{Dt} = \frac{1}{C_{p,S}} \frac{Dq}{Dt} \quad (\text{A.28})$$

The other two coefficients in the temperature equation are:

$$\left. \frac{\partial T}{\partial S} \right|_{\eta,p} = \frac{1}{S \pi_T} \quad (\text{A.29})$$

$$\left. \frac{\partial T}{\partial p} \right|_{\eta, S} = \Gamma(p, \eta, S) = \frac{\gamma - 1}{\rho \alpha c_s^2} = \frac{\alpha T}{\rho C_{p, S}} \quad (\text{A.30})$$

where

$$\begin{aligned} \pi_T = \frac{1}{S} \left. \frac{\partial S}{\partial T} \right|_{\eta, p} & \text{ is the thermohalinity coefficient} \\ \Gamma(p, \eta, S) & \text{ is the adiabatic lapse rate} \\ \gamma = \frac{C_{p, S}}{C_{v, S}} & \text{ is the ratio of specific heats} \\ \alpha = \frac{-1}{\rho} \left. \frac{\partial \rho}{\partial T} \right|_{p, S} & \text{ is the isobaric/isohaline thermal expansion coefficient} \\ c_s & \text{ is the speed of sound, } c_s^2 = \left. \frac{\partial P}{\partial \rho} \right|_{\eta, S} \end{aligned}$$

all of which can be measured.

Substituting all the above coefficients into the temperature equation yields:

$$\frac{DT}{Dt} = \frac{1}{\rho C_{p, S}} (\mathcal{H} - \nabla \cdot \mathbf{F}_q) + \Gamma(p, \eta, S) \frac{Dp}{Dt} + \frac{1}{S \pi_T} \frac{DS}{Dt} \quad (\text{A.31})$$

The thermohalinity is very small and since salinity is conserved the term can be dropped with little need for rigorous justification. Temperature is not conserved due to the compressibility of the fluid. A more convenient quantity is the potential temperature θ , defined as the temperature a parcel would have were it moved adiabatically to some reference level:

$$\theta \equiv T - \int_{p_0}^p \Gamma(p', \eta, S) dp' \quad (\text{A.32})$$

Potential temperature is then conserved except in the presence of external heating:

$$\frac{D\theta}{Dt} = \frac{1}{\rho C_{p, S}} (\mathcal{H} - \nabla \cdot \mathbf{F}_q) \quad (\text{A.33})$$

In the Eulerian form, the temperature equation reads:

$$\frac{\partial}{\partial t} \rho \theta + \nabla \cdot (\rho \theta \mathbf{u}) = \frac{\mathcal{H} - \nabla \cdot \mathbf{F}_q}{C_{p, S}} \quad (\text{A.34})$$

The density of sea water is a function (measured empirically) of potential temperature (θ), salinity (S) and pressure (p). The equation of state is normally written succinctly as:

$$\rho = \rho(\theta, S, p) \quad (\text{A.35})$$

although it is represented by a high-order, empirically determined polynomial.

A.5 Pressure equation (prognostic)

The resulting system is seven equations in seven variables ρ , \mathbf{u} , S , θ , p ; continuity (A.1), momentum (A.12), salt (A.18), temperature (A.34) and the equation of state (A.35). The system is not well suited for practical computations since there is no explicit equation for the pressure. There are two explicit equations for the density; one prognostic (continuity) and one diagnostic (equation of state).

Differentiating the equation of state, the continuity equation can be related to changes in pressure:

$$\frac{D\rho}{Dt} = \left. \frac{\partial \rho}{\partial T} \right|_{S, p} \frac{D\theta}{Dt} + \left. \frac{\partial \rho}{\partial S} \right|_{\theta, p} \frac{DS}{Dt} + \left. \frac{\partial \rho}{\partial p} \right|_{\theta, S} \frac{Dp}{Dt} \quad (\text{A.36})$$

If the fluid is adiabatic, the derivatives of temperature and salt can be eliminated. The total derivative of density can be substituted from the continuity equation yielding:

$$\frac{Dp}{Dt} = -\rho c_s^2 \nabla \cdot \mathbf{u} \quad (\text{A.37})$$

where the speed of sound is given by $c_s^2 = \left. \frac{\partial p}{\partial \rho} \right|_{\theta, S}$. The pressure equation replaces the continuity equation so that there is now one explicit equation for each variable. It would be wrong to replace the equation of state with the pressure equation since that would change the number of prognostic equations. There are six natural modes in the system as described, and currently six prognostic equations.

Appendix B

Scaling of non-hydrostatic effects

To aid the examination of the hydrostatic approximation, the Boussinesq equations of motion (2.11) will be temporarily non-dimensionalised. External parameters will be assumed. The horizontal and vertical length scales of the motion will be L and H . The horizontal and vertical velocity scales will be U and W . A density field will be written as a perturbation, scaled by $\Delta\rho$, from a reference stratification of

$$-\frac{\rho_o}{g}N^2\frac{\partial}{\partial\tilde{z}}\bar{\rho}(\tilde{z})$$

where $N^2 = -(g/\rho_o)\frac{\partial\rho_o}{\partial z}$ is the Brunt-Vasala or buoyancy frequency. Pressure will be split into two parts, a hydrostatic part, scaled by P , and a non-hydrostatic part scaled by rP where r is the ratio of non-hydrostatic pressure to hydrostatic pressure. For the purposes of scaling the horizontal divergence, the non-dimensional horizontal flow is further split into a rotational and a divergent part, the later being ϵ times smaller:

$$\rho' = \frac{\rho_o}{g}N^2H\bar{\rho}(\tilde{z}) + \Delta\rho\tilde{\rho} \quad (\text{B.1})$$

$$p' = P\tilde{p}_H + rP\tilde{p}_{NH} \quad (\text{B.2})$$

$$\{u, v\} = U\{\tilde{u}, \tilde{v}\} \quad (\text{B.3})$$

$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}}_\psi + \epsilon\tilde{\mathbf{u}}_\chi \quad (\text{B.4})$$

$$\nabla_h \cdot \tilde{\mathbf{u}}_\psi = 0 \quad \nabla_h \wedge \tilde{\mathbf{u}}_\chi = 0 \quad (\text{B.5})$$

$$w = W\tilde{w} \quad (\text{B.6})$$

$$\{x, y\} = L\{\tilde{x}, \tilde{y}\} \quad (\text{B.7})$$

$$z = H\tilde{z} \quad (\text{B.8})$$

Non-dimensional variables are indicated by a tilde. The Boussinesq equations of motion (2.11) can then be written:

$$\frac{U^2}{L}\frac{d\tilde{\mathbf{u}}_h}{dt} + 2\Omega U(\sin\phi\hat{k} \wedge \tilde{\mathbf{u}}_h - \frac{W}{U}\mathbf{u}\tilde{w}\cos\phi) + \frac{P}{\rho_o L}(\nabla_h\tilde{p}_H + r\nabla_h\tilde{p}_{NH}) = 0 \quad (\text{B.9})$$

$$\frac{UW}{L}\frac{d\tilde{w}}{dt} + 2\Omega U\tilde{u}\cos\phi + \frac{rP}{\rho_o H}\frac{\partial\tilde{p}_{NH}}{\partial\tilde{z}} = 0 \quad (\text{B.10})$$

$$g\Delta\rho\tilde{\rho} + \frac{P}{H}\frac{\partial\tilde{p}_H}{\partial\tilde{z}} = 0 \quad (\text{B.11})$$

$$\frac{U\Delta\rho}{L}\frac{d\tilde{\rho}}{dt} - W\frac{\rho_o}{g}N^2\frac{\partial\tilde{\rho}}{\partial\tilde{z}}\tilde{w} = 0 \quad (\text{B.12})$$

$$\frac{\epsilon U}{L}\tilde{\nabla}_h \cdot \tilde{\mathbf{u}}_\chi + \frac{W}{H}\frac{\partial\tilde{w}}{\partial\tilde{z}} = 0 \quad (\text{B.13})$$

The last three equations, hydrostatic, buoyancy and continuity, contain two terms each and so the parameters multiplying each pair of terms must balance. Thus, the magnitude of hydrostatic pressure P , density perturbation $\Delta\rho$ and vertical velocity can all be eliminated:

$$P = g\Delta\rho H \quad (\text{B.14})$$

$$\frac{g\Delta\rho}{\rho_o H} = \frac{WL}{UH}N^2 = \epsilon N^2 \quad (\text{B.15})$$

$$W = \epsilon\frac{H}{L}U \quad (\text{B.16})$$

Eliminating these parameters from the equations of motion then gives:

$$R_o \frac{d\tilde{\mathbf{u}}_h}{dt} + (\sin \phi \hat{k} \wedge \tilde{\mathbf{u}}_h - \epsilon \delta \tilde{\mathbf{u}} \tilde{w} \cos \phi) + \epsilon R_o R_i (\nabla_h \tilde{p}_H + r \nabla_h \tilde{p}_{NH}) = 0 \quad (\text{B.17})$$

$$R_o \frac{d\tilde{w}}{dt} + \frac{1}{\epsilon \delta} \tilde{u} \cos \phi + \frac{r R_o R_i}{\delta^2} \frac{\partial \tilde{p}_{NH}}{\partial z} = 0 \quad (\text{B.18})$$

$$\tilde{\rho} + \frac{\partial \tilde{p}_H}{\partial z} = 0 \quad (\text{B.19})$$

$$\frac{d\tilde{p}}{dt} + \frac{\partial \tilde{p}}{\partial z} \tilde{w} = 0 \quad (\text{B.20})$$

$$\tilde{\nabla}_h \cdot \tilde{\mathbf{u}}_\chi + \frac{\partial \tilde{w}}{\partial z} = 0 \quad (\text{B.21})$$

Here, $R_o \equiv \frac{U}{2\Omega L}$ is the Rossby number and $R_i \equiv \frac{N^2 H^2}{U^2}$ is the Richardson number.

A geostrophically adjusted flow will have a zero order balance between the Coriolis term and horizontal hydrostatic pressure gradient: $\epsilon R_o R_i = 1$. Typically, because of the influence of rotation (gradient flow is then Rossby number smaller than rotational flow), $\epsilon = R_o$ and so the quasi-geostrophic scaling is recovered; $R_i R_o^2 = 1$.

Two scalings for the non-hydrostatic parameter r can be obtained from the vertical momentum equation:

- i) $r = \delta^2 / R_i$ At high latitudes where the planetary rotation vector points nearly vertically upwards, the vertical non-hydrostatic pressure gradient can be balanced with the vertical accelerations. In this instance, even at an aspect ratio of one, the non-hydrostatic ratio is typically a small number unless the Richardson number is small, as for an unstratified fluid.
- ii) $r = \delta / (\epsilon R_o R_i)$ At lower latitudes, on the planetary scale, horizontal coriolis effects become more significant than vertical accelerations. Under the quasi-geostrophic scaling above, r then goes like the aspect ratio. This is considerably larger than scaling i) but at the relatively long horizontal scales consistent with QG scaling, the aspect ratio is then a very small number.

The smallness of the parameter r justifies neglect of the non-hydrostatic pressure gradients in both the vertical and horizontal momentum equations. The Coriolis term in the vertical momentum equation (the vertical acceleration is $R_o \epsilon \delta$ smaller than the coriolis term) is δ smaller than the buoyancy term if the flow is geostrophically adjusted ($\epsilon R_o R_i = 1$). This is conventionally the justification behind the hydrostatic approximation since the aspect ratio is inevitably a small number for almost all oceanic applications.

Appendix C

Solution for zonal flow over a Gaussian bump

Huppert [Hup75] solved the problem of stratified flow over an isolated topographic feature. The solution is for an f-plane in an infinite domain. The non-dimensional analytic solution from Huppert is:

$$\psi(r, \theta, \eta) = -r \sin \theta + h_o B R^{-1} \int_0^{\infty} g(t, \eta) \hat{h}(t) J_o(tr) dt \quad (\text{C.1})$$

where the kernel is

$$g(t, \eta) = \frac{\cosh Bt(1 - \eta)}{\sinh Bt} \quad (\text{C.2})$$

and $\hat{h}(t)$ is the zeroth-order Hankel transform of $h(r)$:

$$\hat{h}(t) = \int_0^{\infty} r' h(r') J_o(tr') dr' \quad (\text{C.3})$$

Here, $B = \frac{NH}{fL}$ is the Burger number and $R = \frac{V}{fL}$ is the Rossby number, both indicative of the upstream flow. The length scale L is the inherent length scale of the topographic feature. r is the non-dimensional lateral radius (non-dimensionalised by L) and $\eta = z/H$ is the non-dimensional height.

$h(r')$ is a non-dimensional shape function of maximum value unity. h_o is the non-dimensional height of the topographic anomaly; the fraction of the upstream fluid depth. The dimensional depth of the bottom surface is therefore given by:

$$z_{bottom} = H h_o h(r) \quad (\text{C.4})$$

For the shape function

$$h(r) = e^{-\frac{1}{2}r^2} \quad (\text{C.5})$$

the Hankel transform of $h(r)$ can be evaluated:

$$\begin{aligned} \hat{h}(t) &= \int_0^{\infty} r' h(r') J_o(tr') dr' \\ &= [-h(r') J_o(tr')]_0^{\infty} - t \int_0^{\infty} h(r') J_1(tr') dr' \\ &= 1 - \int_0^{\infty} e^{-\frac{1}{2} \frac{\xi^2}{t^2}} J_1(\xi) d\xi \\ &= 1 - [1 - e^{-\frac{1}{2}t^2}] \\ &= e^{-\frac{1}{2}t^2} \end{aligned} \quad (\text{C.6})$$

where the co-ordinate transform $\xi = tr$ and the relation

$$\int_0^{\infty} e^{-\frac{1}{2}\frac{\xi^2}{t^2}} J_1(\xi) d\xi = 1 - e^{-\frac{1}{2}t^2} \quad (\text{C.7})$$

have been used.

The Taylor column part of the solution is then given by:

$$\int_0^{\infty} \frac{\cosh Bt(1-\eta)}{\sinh Bt} e^{-\frac{1}{2}t^2} J_0(rt) dt \quad (\text{C.8})$$

which is integrated numerically for use in chapter 6.

Bibliography

- [AL77] Akio Arakawa and Vivian R. Lamb. *Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model*, volume 17 of *Methods in Computational Physics*, pages 174–267. Academic Press, 1977.
- [Ape87] J. R. Apel. *Principles of Ocean Physics*, volume 38 of *International Geophysics Series*. Academic Press, 1987.
- [BC68] K. Bryan and M.D. Cox. A nonlinear model of an ocean driven by wind and differential heating: Parts i and ii. *J. Atmos. Sci.*, 25:945–978, 1968.
- [Bry63] Kirk Bryan. A numerical investigation of a nonlinear model of a wind-driven ocean. *J. Atmos. Sci.*, 20:594–606, 1963.
- [Bry69] Kirk Bryan. A numerical method for the study of the circulation of the world ocean. *J. Comp. Phys.*, 4:347–376, 1969.
- [Cox84] M.D. Cox. A primitive equation three-dimensional model of the ocean. Technical report, GFDL, Princeton Univ., 1984.
- [DD92] John K. Dukowicz and Arkady S. Dvinsky. Approximate factorization as a high order splitting for the implicit incompressible flow equations. *J. Comp. Phys.*, 102(2):336–347, 1992.
- [Dea71] J.W. Deardorff. On the magnitude of the subgrid scale eddy coefficient. *J. Comp. Phys.*, 7:120–133, 1971.
- [DS94] John K Dukowicz and Richard D Smith. Implicit free-surface method for the Bryan-Cox-Semtner ocean model. *J. Geophys. Res.*, 99(C4):7991, 8014 1994.
- [DSM93] John Dukowicz, Richard Smith, and R.C. Malone. A reformulation and implementation of the Bryan-Cox-Semtner ocean model on the connection machine. *J. Ocean. Atmos. Tech.*, 10(2):195–208, 1993.
- [FR91] Michael S. Fox-Rabinovitz. Computational dispersion of horizontal staggered grids for atmospheric and ocean models. *Mon. Wea. Rev.*, 119:1624–1639, July 1991.
- [GB79] Michael Ghil and Ramesh Balgovind. A fast Cauchy-Riemann solver. *Mathematics of Computation*, 33(146):585–635, 1979.
- [Han71] R.L. Haney. *A numerical study of the large-scale response of an ocean circulation to surface heat and momentum flux*. PhD thesis, UCLA, 1971.
- [Han84] Y.-J. Han. A numerical world ocean general circulation model. *DAO*, 8:141–172, 1984.
- [Hup75] H.E. Huppert. Some remarks on the initiation of inertial Taylor columns. *J. Fluid. Mech.*, 67:397–412, 1975.
- [HW80] George J. Haltiner and Roger Terry Williams. *Numerical Prediction and Dynamic Meteorology*. John Wiley & Sons, 1980.
- [HWY91] D. B. Haidvogel, J.L. Wilkin, and R. Young. A semi-spectral primitive equation ocean circulation model using sigma and orthonogonal curvilinear coordinates. *J. Comp. Phys.*, 94:151–185, 1991.
- [Jr.86] A.J. Semtner Jr. *History and Methodology of Modelling the Circulation of the World Ocean*, pages 23–32. Advanced Physical Oceanographic Numerical Modelling. J.J. O’Brien, 1986.

- [Kim79] J.-W. Kim. Design and preliminary performance of the OSU four-level oceanic general circulation model. Technical report, Oregon State Univ., 1979.
- [Lev82] S. Levitus. Climatological atlas of the world ocean. *NOAA*, Professional Paper 13, 1982.
- [LJVPO] Sonya A. Legg, Helen Jones, and Martin Visbeck. The role of baroclinic instability in open ocean deep convection. *in preparation*, JPO.
- [MA76] F. Mesinger and A. Arakawa. Numerical methods used in atmospheric models. Technical report, WMO/ICSU Joint Organizing Committee GARP Publ. Series, 1976.
- [Mel92] G. L. Mellor. *Program in Atmospheric and Oceanographic Sciences Report*, chapter User's guide for a three-dimensional, primitive equation, numerical ocean model. Princeton Univ., Princeton, N.J., 1992.
- [Mos66] Moskau, editor. *Numerical computations of the wind-induced mass transport in a stratified ocean*, volume Proc. Symp. Math.-Hydrodyn. Invest. Phys. Processes in the Sea, 1966.
- [MRMH93] E. Maier-Reimer, U. Mikolajewicz, and K. Hasselmann. Mean circulation of the Hamburg LSG OGCM and its sensitivity to the thermohaline surface forcing. *J. Phys. Oceanogr.*, 23:731–757, 1993.
- [Obe88] J.M Oberhuber. An atlas based on the 'COADS' data set: The budgets of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. Technical Report Report 15, Max-Planck-Institute for Meteorology, 2000 Hamburg 13, Bundesstrasse, 1988.
- [Ped79] Joseph Pedlosky. *Geophysical Fluid Dynamics*. Springer-Verlag New York Inc., 1979.
- [Sar55] Artem Sarkisyan. (title unknown). *Izv. Akad. Nauk SSSR*, 6(554), 1955. see Semtner 1986.
- [SBB90] Linda T. Smith, Douglas B. Boudra, and Rainer Bleck. A Wind-Driven Isopycnal Coordinate Model of the North and Equatorial Atlantic Ocean. 2. The Atlantic-Basin Experiments. *J. Geophys. Res.*, 95(C8):13,105–13,128, August 1990.
- [SC88] A.J. Semtner and R. Chervin. A simulation of the global ocean circulation with resolved eddies. *J. Geophys. Res.*, 8:735–754, 1988.
- [Sem74] A.J. Semtner. An oceanic general circulation model with bottom topography. Numerical simulation of Weather and Climate 9, UCLA, 1974.
- [SM77] Albert J. Semtner, Jr. and Yale Mintz. Numerical simulation of the gulf stream and mid-ocean eddies. *J. Phys. Oceanogr.*, 77:208–230, 1977.
- [Sma63] J.S. Smagorinsky. General circulation experiments with the primitive equations. i: The basic experiment. *Mon. Wea. Rev.*, 91:99–164, 1963.
- [Tak74] K. Takano. A general circulation model for the world ocean. Technical Report 8, Univ. California, Los Angeles, 1974.
- [Tak75] K. Takano. *A numerical simulation of the world ocean circulation: preliminary results*, pages 121–129. Numerical Models of Ocean Circulation. Acad. of Sciences, Washington, D.C., 1975.
- [TOL89] K.E. Trenberth, J.G. Olson, and W.G. Large. A global ocean wind stress climatology based on ECWMF analyses. Technical report, NCAR/TN-388+STR, NCAR, Colorado, 1989.
- [Waj86] Roxana C. Wajsowicz. Free planetary waves in finite-difference numerical models. *J. Phys. Oceanogr.*, 16:773–789, April 1986.
- [Xu94] Weimin Xu. *A C-grid Ocean Circulation Model and Eddy Simulation*. PhD thesis, McGill University, Montreal, Quebec, March 1994.